

FILOGENIA MOLECULAR DE GENES DA FAMÍLIA MADS BOX  
RESPONSÁVEIS PELO DESENVOLVIMENTO DOS ÓRGÃOS  
FLORAIS DE MAMOEIRO (*Carica papaya* L.).

**MILENE DE FIGUEIREDO**

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE  
DARCY RIBEIRO – UENF

CAMPOS DOS GOYTACAZES-RJ  
MARÇO-2016

FILOGENIA MOLECULAR DE GENES DA FAMÍLIA MADS BOX  
RESPONSÁVEIS PELO DESENVOLVIMENTO DOS ÓRGÃOS  
FLORAIS DE MAMOEIRO (*Carica papaya* L.).

**MILENE DE FIGUEIREDO**

“Tese apresentada ao Centro de Ciências e  
Tecnologias Agropecuárias da Universidade  
Estadual do Norte Fluminense Darcy Ribeiro,  
como parte das exigências para obtenção do  
título de Doutora em Genética e Melhoramento  
de Plantas.”

Orientadora: Prof<sup>a</sup>. Telma Nair Santana Pereira

CAMPOS DOS GOYTACAZES – RJ  
MARÇO – 2016

FILOGENIA MOLECULAR DE GENES DA FAMÍLIA MADS BOX  
RESPONSÁVEIS PELO DESENVOLVIMENTO DOS ÓRGÃOS  
FLORAIS DE MAMOEIRO (*Carica papaya* L.).

**MILENE DE FIGUEIREDO**

“Tese apresentada ao Centro de Ciências e Tecnologias Agropecuárias da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Doutora em Genética e Melhoramento de Plantas.”

Aprovada em 29 de Março de 2016

Comissão Examinadora:

---

Prof. Messias Gonzaga Pereira (PhD., Melhoramento de Plantas) - UENF

---

Prof. Pedro Corrêa Damasceno Junior (D.Sc., Genética e Melhoramento de Plantas) - UFRRJ

---

Dr<sup>a</sup>. Fernanda Abreu Santana Arêdes (D.Sc., Genética e Melhoramento de Plantas) - UENF

---

Prof<sup>a</sup>. Telma Nair Santana Pereira (PhD., Melhoramento de Plantas) - UENF  
(Orientadora)

## **AGRADECIMENTO**

À Universidade Estadual Norte Fluminense e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, pela oportunidade de Estudo em uma Instituição de Ensino Superior de alta qualidade.

À Dr<sup>a</sup> Telma Nair Santana Pereira, minha orientadora, pela oportunidade que me ofereceu, pelos auxílios, conversas, conselhos, amizade e por confiar e acreditar no meu trabalho.

Ao prof. Dr. Messias Pereira, pela excelência do ensino dado para a elaboração dos experimentos, por abrir as portas do laboratório para o desenvolvimento da parte molecular do trabalho.

À Banca examinadora- Dr. Pedro Damasceno e Dra. Fernanda Abreu Santana- por aceitarem o convite e por terem contribuído tanto na estruturação da dissertação.

À Fundação de Amparo à Pesquisa do Rio de Janeiro (Faperj), à Coordenadoria de Aperfeiçoamento do Pessoal de Nivel Superior (Capes), pela oportunidade do desenvolvimento da pesquisa e pela concessão de bolsa.

À empresa CALIMAN Agrícola, pelo interesse, apoio e infraestrutura, em especial, ao diretor técnico Geraldo A. Ferreguetti, e aos técnicos da fazenda.

A todos os professores da Genética e Melhoramento de Plantas da UENF que contribuíram para com o meu título de doutora.

Aos colegas do Laboratório de Citogenética da UENF, pelo auxílio e pela amizade.

Aos amigos do Laboratório de Marcadores Moleculares da UENF, em especial Vitorinha, Marcelinha e Fernandinha! Fui privilegiada por ter vocês comigo nessa caminhada.

Ao querido José Daniel, secretário da pós-graduação, pela inestimável e constante ajuda, principalmente por sempre me alertar das tarefas de que me esquecia.

Às amigas da UFMG, Mel e Cassi, que foram essenciais para a elaboração dos experimentos.

À minha grande amiga Liliam Candido, pela ajuda na procura de um professor na área de Bioinformática e pelo auxílio e teto no desenvolvimento da tese. Só tenho a agradecer.

Ao Professor Dr. Rodrigo Pereira, da Universidade Federal da Grande Dourados, por me proporcionar um aprendizado excelente em uma área fantástica da biologia: a bioinformática.

À minha família, pelos grandes momentos e por sempre permanecer ao meu lado nas minhas inúmeras decisões e indecisões, em especial, à minha querida mãe, por sempre acreditar que meus passos seriam vitoriosos.

Ao meu amigo e namorado Leonardo Santos, pela paciência na espera, aguentando firme a distância e sempre me ajudando e apoiando. Te amo sempre!

A Deus, pela vida. Obrigada pela oportunidade de viver e pelo auxílio nas minhas escolhas.

## SUMÁRIO

RESUMO .....	vi
ABSTRACT .....	viii
1. INTRODUÇÃO .....	1
2. OBJETIVOS .....	4
3. REVISÃO DE LITERATURA .....	6
3.1. Importâncias Socioeconômicas.....	6
3.2. A Filogenética da Família Caricaceae.....	8
3.3. O gênero Carica .....	10
3.4. Cromossomos sexuais e biologia floral da família Caricaceae .....	11
3.5. Genes MADS-Box em plantas .....	12
3.6. Redes reguladoras de genes no desenvolvimento floral.....	15
3.7. Bioinformática e Análises Filogenéticas no Melhoramento de Plantas	16
4. MATERIAL E MÉTODOS .....	19
4.1. Material Genético .....	19
4.2. Extração do DNA genômico .....	20
4.3. Seleção de genes homeóticos e desenvolvimento de primers.....	20
4.4. Reação em cadeia da polimerase - PCR .....	21
4.5. Análises das sequências.....	21
4.6. Análises Filogenéticas moleculares .....	21
5. RESULTADOS E DISCUSSÃO .....	28
5.1. Coleta do material vegetal e extração do DNA genômico .....	23

5.2. Iniciadores utilizados nas análises .....	23
5.3. Reações em cadeia da Polimerase PCR .....	25
5.4. Sequenciamento e análises de bioinformática .....	26
5.5. Filogenia molecular dos genes FUL e SOC1 em mamoeiro .....	37
6.CONCLUSÕES E PERSPECTIVAS FUTURAS .....	46
REFERÊNCIAS BIBLIOGRÁFICAS.....	49

## RESUMO

FIGUEIREDO, Milene, DSc., Universidade Estadual do Norte Fluminense Darcy Ribeiro, Março de 2016, Filogenia molecular de genes da família MADS Box responsáveis pelo desenvolvimento dos órgãos florais de mamoeiro (*Carica papaya* L.). Orientadora: Telma Nair Santana Pereira. Conselheiros: Messias Gonzaga Pereira e Gonçalo Apolinário de Souza Filho.

O mamoeiro cultivado comercialmente (*Carica papaya* L.) pertence à família Caricaceae, está dividida em seis gêneros. A importância econômica do gênero *Carica* reside principalmente na produção de frutos por parte de sua única espécie, *Carica papaya*, amplamente cultivada na região dos trópicos. O desenvolvimento floral é controlado por uma rede complexa de interações entre fatores de transcrição, a maioria deles pertencente à família MADS-box. Para entender a evolução e mecanismos regulatórios da determinação sexual do mamoeiro, a identificação dos genes associados aos cromossomos sexuais é de extrema relevância. Pela importância desse fato, foram selecionados os genes de identidade floral *Suppressor of over expression of Constans 1- SOC1* e *Fruitfull-FUL* com o objetivo de realizar uma análise de filogenia molecular em genótipos de mamoeiro que apresentam uma importância comercial no mercado nacional e internacional. Para a extração do DNA genômico, utilizou-se folhas de plantas hermafroditas de quatro genótipos de mamoeiro *Sunrise Solo 72/12*, *JS12*, *Golden* e híbrido *UENF-Caliman* e folhas de plantas femininas do genótipo *Cariflora*. Os iniciadores para amplificação da região específica do genoma e posterior sequenciamento, foram construídos utilizando como ferramenta principal

o *PRIMER3 plus*. Para o sequenciamento dos fragmentos de PCR, utilizou-se o equipamento *ABI-Prism 3500 Genetic Analyzer* e para a análise da qualidade das sequências utilizou-se o *BioEdit Sequence Alignment Editor 7.1.9* e o *CodonCode Aligner* versão 6.0.2 o que resultou na obtenção de *contigs* dos segmentos gênicos. Para o alinhamento global dos *contigs* e dos aminoácidos, utilizou-se o programa *Clustal X 2.1* com iteração em cada etapa de alinhamento, tendo as árvores filogenéticas sido geradas através do programa *MEGA Beta 6* e o *CLC 7.5*. Os domínios MADS-box de todos os 10 *contigs* formados foram identificados com o auxílio do banco de dados biológicos on-line *Pfam*. As sequências de aminoácidos foram comparadas com o banco de dados *Uniprot* e *Swiss-Prot* e apresentaram uma alta porcentagem de similaridade com os genes MADS-box de outras espécies. Todas as sequências dos aminoácidos traduzidas dos genes FUL e SOC1 apresentaram aproximadamente 60 resíduos, incluindo a Metionina inicial (ATG- códon iniciador). Para o gene FUL, o genótipo Cariflora não apresentou diferença estrutural com a sequência de aminoácidos depositada no *Phytozome*; já os genótipos JS12, SS72/12, Golden e Calimosa apresentaram substituições conservativas e não conservativas nas sequências dos aminoácidos. Para o gene SOC1, o genótipo SS72/12 não apresentou diferença estrutural quando comparado com a sequência de aminoácidos depositada no *Phytozome*. Os genótipos JS12, Cariflora, Golden e Calimosa apresentaram substituições conservativas e não conservativas nas sequências dos aminoácidos. As sequências dos genes MADS-box utilizadas para a inferência filogenética nesse trabalho têm somente um éxon, o domínio SRF, e foram classificados como sendo genes MADS-box do tipo I. Para a inferência funcional dos genes Tipo I de mamoeiro presentes neste trabalho, análises de expressão gênica precisam ser integradas. Os experimentos de bioinformática irão auxiliar o direcionamento dos experimentos de bancada, os resultados obtidos podem ser futuramente validados via qPCR.

## ABSTRACT

FIGUEIREDO, Milene, DSc., Universidade Estadual Norte Fluminense Darcy Ribeiro, March of 2016, Molecular phylogeny of Mad-Box family genes, responsible for the development of floral organs of the papaya (*Carica papaya* L.). Adviser: Telma Nair Santana Pereira. Co-Comitee Members: Messias Gonzaga Pereira and Gonçalo Apolinário de Souza Filho.

Papaya commercially grown (*Carica papaya* L.) belongs to the family Caricaceae which is divided into 6 genera. The economic importance of *Carica* genus occupies mainly the production of fruit by the single species, *C. papaya*, widely cultivated in the tropics. Floral development is controlled by a complex network of interactions between transcription factors, most of them belonging to the MADS - box family. To understand the evolution and regulatory mechanisms of sex determination of papaya, the identification of genes associated with sex chromosomes is extremely important. Due to the importance of this fact, there were selected the floral identity genes Suppressor of over expression of Constans 1- SOC1e Fruitfull- FUL, and for phylogenetic analysis. There were used the leaves of plants and hermaphroditic papaya genotypes 4 - Sunrise Solo 72/12, JS12, Golden and hybrid UENF-Caliman - and leaves of female plants Cariflora genotype for the extraction of genomic DNA. Primers for amplification sequence analysis of the specific region of the genome and subsequent sequencing were constructed using as the main tool Primer3 plus. For sequencing of PCR fragments, it was used the equipment ABI-Prism 3500 Genetic Analyzer. After sequencing, it was verified the quality of the sequences in BioEdit Sequence

Alignment Editor 7.1.9 and obtaining the contigs used the program CodonCodeAligner version 6.0.2. For the global alignment of contigs and amino acids, it was used the Clustal X program 2.1 interaction with each alignment step and phylogenetic trees were generated using the Beta, Mega CLC 6 and 7.5. MADS -box domains of all 10 contigs were identified with the help of biological online database Pfam. The amino acid sequences were compared with the database and UniProt Swiss-Prot and showed a high percentage of similarity with genes from other species. All sequences of amino acids translated from the FUL eSOC1 gene showed approximately 60 residues, including the initial Methionine (ATG- initiator codon). For the FUL gene, Cariflora genotype didn't present structural difference with the amino acid sequence deposited on Phytozome; already the JS12 genotypes SS72/12, Golden and Calimosa presented conservative and non- conservative substitutions in the sequences of amino acids. For SOC1 gene, SS72/12 genotype showed no structural difference with the amino acid sequence deposited on Phytozome. Genotypes OS JS12, Cariflora, Golden and Calimosa showed substitutions in sequences of conservative and non-conservative amino acids. The sequences of the MADS -box genes used for phylogenetic inference in this paper has only one exon. For only present the SRF domain they were classified as MADS -box genes type I. For Functional inference of gene Type I from papaya present in this study, gene expression analysis need to be integrated. Bioinformatics experiments will help the direction of bench experiments and can validate by PCR, the results obtained by bioinformatics.

## 1. INTRODUÇÃO

O mamoeiro (*Carica papaya*) é um membro da família Caricaceae, que é composta de 35 espécies divididas em seis gêneros (Badillo 2000; Carvalho e Renner 2012; Gschwend et al., 2013). Os membros desta família são conhecidos por apresentarem dimorfismo sexual; 32 espécies são dióicas, uma espécie é monóica, e duas espécies, incluindo mamão, são trióicas. O gênero *Carica* é diplóide com número básico de cromossomos haplóides igual a 9 (Storey 1941 citado por Dantas et al., 2002). A análise filogenética revelou que dioicia é o estado ancestral da família, e as plantas hermafroditas de mamoeiro evoluíram como uma mutação natural de plantas masculinas, provavelmente selecionadas por seres humanos por ser considerado fenótipo favorável (Carvalho e Renner 2012).

Para entender a evolução e mecanismos reguladores da determinação sexual do mamoeiro, a identificação dos genes associados aos cromossomos sexuais é de extrema importância. A recombinação genética é suprimida em uma região específica entre os cromossomos X e Y ou  $Y^h$ , e estes cromossomos são previstos serem derivados de um par de autossomas (Charlesworth e Charlesworth, 1978; Ming et al., 2011). Esta região de recombinação suprimida no cromossomo Y ou  $Y^h$  está prevista para corresponder à região de determinação sexual, designada como uma região específica para o sexo masculino no cromossomo Y (MSY), ou uma região específica para o hermafroditismo no cromossomo  $Y^h$  (HSY) (Ming et al., 2007). Embora seja claro que as diferenças

estruturais nas regiões MSY e HSY entre os três cromossomos sexuais (X,Y Y<sup>h</sup>) sejam responsáveis pela determinação do sexo em mamoeiro, os genes envolvidos na determinação do sexo ainda não foram identificados.

O controle genético do desenvolvimento reprodutivo das plantas tem sido amplamente caracterizado em várias espécies ao longo dos últimos anos, tornando este processo de desenvolvimento um eficiente sistema para observar a evolução ao nível individual dos genes. O desenvolvimento floral é controlado por uma rede complexa de interações entre fatores de transcrição, a maioria deles pertencente à família MADS-box (Schwarz-Sommer et al., 1990;. Mondragon-Palomino e Trontin, 2011).

Duplicação gênica gera cópias redundantes de genes que pode evoluir para diferentes direções através de mutações em sua sequência codificante e/ou regulatória. A presença de genes duplicados adicionais libera a restrição de manter a função original, porque esse papel pode ser garantido pela presença de outro gene duplicado ou particionado entre os genes duplicados. As análises filogenéticas mostram claramente que o aumento dos genes MADS-box, visto em plantas superiores com flores, tem suas origens na duplicação de genes (Theissen et al., 1996; Parenicova et al., 2003). A história da duplicação gênica da família MADS-box tem sido reconstruída por extensivas análises filogenéticas, e a ocorrência de eventos chave de duplicação próxima à origem das angiospermas tem ainda implicado a evolução e a diversificação das angiospermas (Kim et al., 2004; Zahn et al., 2006; Soltis et al., 2007). As duplicações que deram origem à maioria dos genes da família MADS-box que controlam o desenvolvimento da flor podem ser, em grande parte, atribuídas a vários eventos de duplicação do genoma inteiro, que têm sido uma característica regular da evolução da planta (Causier et al., 2005; Zahn et al., 2005). Em contraste, as duplicações que deram origem, muitas vezes, à mais numerosa classe I genes MADS-box são geralmente associadas a eventos de duplicação locais de menor escala (Nam et al., 2004).

Assim, no presente estudo, as sequências gênicas de dois genes da família MADS BOX, *Supressor of Over expression of Constans 1- SOC1* e *Fruitfull- FUL*, importantes no desenvolvimento floral de plantas, foram sequenciadas nos genótipos de mamoeiro Cariflora, Golden, JS12, SS72/12 e Calimosa. O objetivo dessa análise, utilizando uma abordagem filogenética, foi estudar a evolução dos

genes MADS-Box tipo I em mamoeiro e sua associação funcional, utilizando o banco de dados Uniprot e Swissprot com diversas espécies.

## 2. OBJETIVOS

### 2.1. Objetivo Geral

Esta pesquisa teve por objetivo fazer a análise filogenética dos genes *Supressor of Over expression of Constans 1- SOC1* e *Fruitfull- FUL*, pertencentes à família MADS-BOX, dos genótipos de mamoeiro do banco de germoplasma da UENF/Caliman<sup>®</sup>.

### 2.2. Objetivos Específicos

- Selecionar genes homeóticos da família MADS- BOX no genoma do mamoeiro e construir iniciadores específicos para os genes selecionados;
- Selecionar e coletar os cinco genótipos de mamoeiro - Cariflora, Golden, JS12, SS72/12 e o híbrido UENF/Caliman 01- na área de plantio comercial da empresa Caliman Agrícola S/A, localizada no município de Sooretama/ES;
- Extrair o DNA genômico dos cinco genótipos selecionados;
- Amplificar a sequência gênica dos genes SOC1 e FUL no genoma dos cinco genótipos do mamoeiro;

- Sequenciar os produtos das ampliações dos genes SOC1 e FUL;e
- Fazer as análises das sequências geradas e análises filogenéticas moleculares.

### 3. REVISÃO DE LITERATURA

#### 3.1. Importâncias Socioeconômicas

A produção mundial de mamão atingiu 12,5 milhões de toneladas em 2013, tendo como principais produtores a Índia, Brasil, Indonésia, Nigéria e México. O Brasil, segundo maior produtor, respondeu com 12,6% da produção mundial, no ano de 2013, produzindo 1,5 milhões de toneladas da fruta, o que dá aproximadamente 49,5 mil kg/ha. O maior produtor foi a Índia, com 5,5 milhões de toneladas, o que representou 44,1% da produção mundial (FAOSTAT, 2015).

O volume mundial de mamão exportado em 2012 foi de 271,8 mil toneladas, correspondendo a US\$ 209,4 milhões. O México, apesar de ser o sexto país maior produtor de mamão, aparece em primeiro lugar como exportador, sendo responsável por aproximadamente 40% das exportações. O Brasil ocupa a segunda posição, com 9,6% da exportação mundial, seguido da Guatemala, com 8,2% (FAOSTAT, 2015).

O mamoeiro, apesar de ser cultivado em praticamente todo o território nacional, é na Bahia e no Espírito Santo que está concentrada a maior produção; juntos, esses dois estados respondem com 71% da produção brasileira (IBGE, 2015).

A exportação brasileira em 2014 cresceu 17,9% (33,1 mil t), sendo o Espírito Santo responsável por 30,4% dessas exportações. O Rio grande do Norte aparece como segundo maior exportador, sendo responsável por 26,1% das exportações até agosto de 2015, superando a Bahia, que exportou 6.966 toneladas, com pequena elevação sobre 2013 (MDIC, 2015).

Os países maiores importadores são Estados Unidos e Cingapura, juntos, representam 63% das importações mundiais da fruta (FAOSTAT, 2015). Portugal é o país maior importador de mamão brasileiro, seguido da Espanha, Holanda, Reino Unido, Alemanha e Estados Unidos (MDIC, 2015).

As cultivares de mamoeiro mais exploradas no Brasil são classificadas em dois grupos conforme o tipo de fruto: Solo e Formosa. As variedades do grupo Formosa são adequadas somente à comercialização no mercado interno, enquanto as do grupo Solo são comercializadas nos mercados interno e externo (Dantas e Castro Neto, 2000). A empresa Caliman tem duas unidades de produção, uma no Norte do Espírito Santo, na cidade de Linhares com uma área total de 480 ha, e uma unidade no Rio Grande do Norte, nas cidades de Pureza (Agreste) e Jandaíra (Semiárido) com área de 150 ha, totalizando 630 ha de área total. As variedades plantadas pela Caliman são a Golden, THB, Aliança, os híbridos Calimosa e Tainung 01 e outras cultivares híbridas em desenvolvimento, não fugindo à média nacional (Toda Fruta, 2015).

Em uma lista de importância nutricional com 38 frutos considerados comuns, o mamão está na primeira posição, tendo como referência a porcentagem de ingestão diária recomendada nos Estados Unidos para a Vitamina A, C, potássio, ácido fólico, niacina, tiamina, riboflavina, ferro, cálcio e fibra (Ming et al., 2008).

A papaína é uma importante enzima proteolítica, encontrada no látex dos frutos verdes, utilizada na indústria têxtil, farmacêutica e de cosméticos, que pode ser apresentada nas formas refinada, líquida ou purificada. Em termos de evolução, a papaína pode estar associada à proteção ao ataque de predadores herbívoros e frutívoros (El Moussaoui et al., 2001). A indústria da cerveja é a principal consumidora de papaína, com cerca de 75% da demanda, sendo utilizada na clarificação e estabilização da cerveja.

Muitos compostos químicos biologicamente ativos têm sido isolados do mamoeiro e estudados quanto à sua ação. Essas atividades biológicas são relatadas em extratos brutos ou em diferentes frações de sementes, folhas, raízes, caule cascas e látex e têm sido utilizados na medicina tradicional para o tratamento de várias doenças. Têm sido relatadas, com base na pesquisa científica, várias propriedades terapêuticas e ações farmacológicas do mamoeiro, como propriedades, antimicrobianas (Calzada et al., 2007; Osato et al., 1993),

anti-helmínticas (Panse e Paranjpe, 1943; Krishnakumari e Majumder, 1960; Bose et al., 1961), antimaláricas (Bhat e Surolia, 2001), anti-fertilidade masculina e feminina (Verma e Chinoy, 2001; Udoh et al., 2005; Kapoor et al., 1974; Bodhankar et al., 1974; Adebivi et al., 2002), contra a dengue (Sarala e Paknikar, 2014) e propriedades bioativas específicas com efeitos anticancerígenos (Nguyen et al., 2013).

Há uma quantidade significativa de trabalhos sobre a atividade biológica e possível aplicação destes compostos, portanto, ainda se faz necessária uma extensa investigação sobre sua farmacodinâmica, cinética e normalização, compreendendo ensaios clínicos adequados, explorando sua utilidade terapêutica para combater várias doenças (Krishna et al., 2008).

### 3.2. A filogenia da família Caricaceae

A filogenia molecular é a base para a compreensão da diversidade das espécies e alterações ligadas às taxas de diversificação com eventos históricos, como, por exemplo, a reestruturação das vias de desenvolvimento, duplicação do genoma, ou a dispersão para um novo continente (Mark et al., 2010). De acordo com o *Angiosperm Phylogeny Group*, a ordem *Brassicales*, grupo monofilético pertencente ao clado *Rosids*, é composta por 14 famílias distribuídas em regiões tropicais e subtropicais do mundo (Stevens, 2001; Edger et al., 2015)( Figura 1).

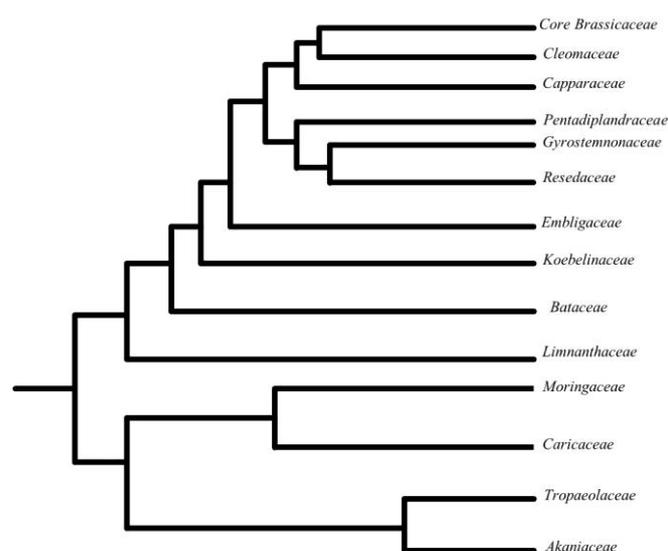


Figura 1: Ordem Brassicales, grupo monofilético composto por 14 famílias. (*Angiosperm Phylogeny Group Adaptado, 2016*).

Entre as famílias da ordem *Brassicales*, temos a família *Brassicaceae* e seu clado irmão *Caricaceae*. A família *Brassicaceae* compreende mais de 3.600 espécies, incluindo *Arabidopsis thaliana*, a primeira espécie a ter seu genoma totalmente sequenciado, sendo um organismo modelo de plantas, incluindo a genética do desenvolvimento, ritmo circadiano e muitos outros aspectos da vida da planta (*The Arabidopsis Genome Initiative* 2000; Müller e Grossniklaus, 2010). A família *Caricaceae* é uma pequena família de plantas, que compreende cerca de seis gêneros e 35 espécies (Figura 2).

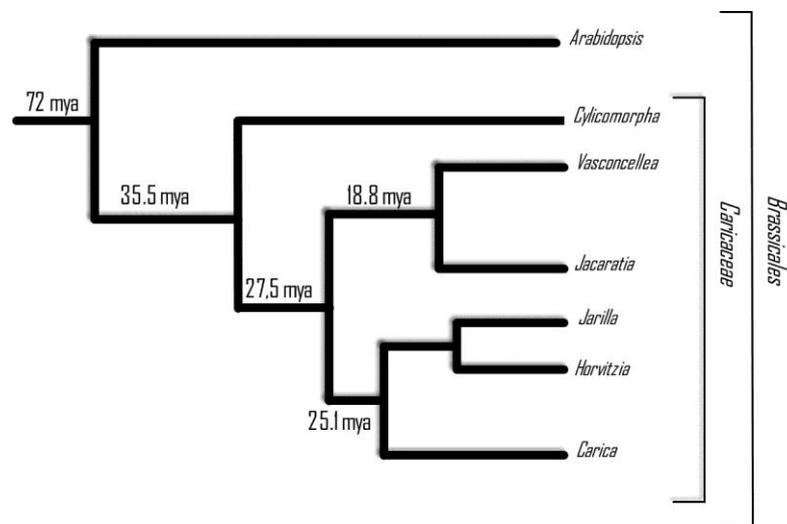


Figura 2: Família *Caricaceae* compartilha um ancestral comum com a família *Brassicaceae*, aproximadamente há 72 milhões de anos.

A filogenia da família *Caricaceae* indica que sua origem foi na África e sua introdução nas regiões neotropicais foi por dispersão a longa distância há cerca de 35,5 milhões de anos, tudo indicando que a família *Caricaceae* atingiu a América do Sul e a América Central há cerca de 27 milhões de anos (Badillo, 1971). Nas regiões neotropicais, a família *Caricaceae* começou a se diversificar sendo que nessa região ainda ocorrem todos os gêneros. A família *Caricaceae* tem uma distribuição anfiatlântica, com duas espécies na África Tropical e 33 espécies na América Central e América do Sul.

*Carica papaya*, a única espécie do gênero *Carica*, é natural da região Neotropical, em um limite de alcance que vai do norte da Flórida ao sul do Paraguai (Badillo, 1971). A forma silvestre do mamoeiro tem frutos muito menores e arredondados (máximo de 7 cm de diâmetro quando maduros) com mesocarpo

muito mais fino (menos de 1 cm) do que a forma cultivada. As espécies silvestres foram encontradas somente nas planícies da América Central a partir de Yucatán, no México, ao sul de Belize, Guatemala oriental e Costa Rica (Coppens d'Eeckenbrugge et al., 2007).

### 3.3 O gênero *Carica*

O gênero *Carica* tem somente uma espécie, *Carica papaya* L., que por motivo de alguns fatores como ampla distribuição pelos espanhóis, grande capacidade de adaptação às condições ambientais nas regiões tropicais e subtropicais e por não existir nenhuma evidência arqueológica direta disponível, é bastante difícil definir uma origem precisa para a espécie (Badillo, 2000; Fuentes e Santamaría, 2014). A região Neotropical é o centro de origem de importantes culturas tropicais, e tem sido sugerido como sendo uma boa candidata para ser também o centro de origem do mamoeiro (Harlan, 1971).

Várias populações silvestres de mamoeiro foram observadas em locais isolados e despovoados no sul do estado de *Yucatán*, México, sul de Belize e Guatemala oriental. Em *Yucatán*, a maior parte das áreas de plantio do mamoeiro, predomina a cultivar Maradol, importada de Cuba, que tem características morfológicas muito distintas dos mamoeiros das populações encontradas ao sul do estado (Coppens d'Eeckenbrugge et al., 2007; Fuentes e Santamaría, 2014).

Em áreas classificadas como sendo centro de origem das espécies, parentes silvestres fornecem alimentos para as comunidades sendo uma importante fonte de renda, pois eles têm genes úteis para aumentar a produtividade e melhorar a qualidade nutricional das frutas (IPGRI 1993).

É difícil indicar o local onde o mamoeiro foi domesticado (Fuentes e Santamaría 2014). Vários estudos incluem o mamoeiro na lista de plantas que os povos Maias teriam usado em seus sistemas agrícolas bem desenvolvidos (de Oviedo, 1959; Dunning et al., 1998; Terán e Rasmussen 1995; Colunga-García Marín e Zizumbo-Villarreal 2004), indicando que papaia (ou "Puut", "colocar" em Maia) foi cultivado no México e Belize, antes da chegada dos espanhóis. Certamente, os Maias da planície tornaram-se uma civilização muito bem-sucedida, sendo uma sociedade agrícola, aparentemente, desde pelo menos 1300 a. C. (Pope et al., 2001).

A introdução da forma silvestre do mamoeiro no Brasil aconteceu na costa nordestina, por intermédio dos portugueses, no final do século XVI, de onde eles o difundiram para a África e, juntamente com os espanhóis, para a Índia (Ctenas et al., 2000). Na década de 1970, a cultura do mamoeiro no Brasil, bastante restrita a pomares domésticos e a pequenas propriedades, sofreu um impacto marcante pela introdução das variedades do grupo Solo, originadas do Havaí (EUA), com frutos pequenos, um sabor mais doce e com polpas mais alaranjadas, diferentes dos frutos amarelos e de sabor insípido, pouco apreciado. Essa introdução ocorreu inicialmente no Estado do Pará, e esses frutos se tornaram conhecidos rapidamente em todo o País como papaia, mamão papaia, mamão do Amazonas, mamão do Pará ou mamãozinho (Dornelles, 2008).

### **3.4. Cromossomo sexual e biologia floral da família Caricaceae**

Dos nove pares de cromossomos do mamoeiro, sete são metacêntricos e dois submetacêntricos (Damasceno Junior et al., 2010; Ming et al., 2008). Os cromossomos são pequenos com morfologia uniforme, tornando difícil sua diferenciação. Nos primeiros estudos de determinação sexual, essa uniformidade gerou dificuldades na identificação dos cromossomos sexuais (Wai et al., 2010). Em uma investigação inicial, observou-se uma separação precoce entre um par de cromossomos durante a anáfase I da meiose de uma célula mãe de grão de pólen em plantas masculinas e hermafroditas (Kumar et al., 1945).

O sexo das plantas de mamoeiro é geneticamente controlado por uma região ligada ao sexo, que se comporta como um cromossomo sexual XY (Liu et al., 2004), mas a expressão fenotípica de sexo em mamoeiro é influenciada por fatores ambientais, incluindo a temperatura, estado nutricional e umidade (Awada e Ikeda, 1957; Awada, 1958). A masculinidade e o hermafroditismo são controlados por regiões do cromossomo Y, que apresenta poucas diferenças em função do sexo; em plantas hermafroditas, utiliza-se a nomenclatura  $Y^h$  ou HSY, e em plantas masculinas, o Y ou MSY. A região correspondente ao cromossomo X é de apenas 3,5 Mb, e ambas as regiões HSY e MSY são de aproximadamente 8,1 Mb. A recombinação das regiões Y e  $Y^h$  com o cromossomo X é suprimida para que regiões específicas do sexo possam ser definidas (Wang et al., 2012).

Qualquer combinação dos cromossomos Y e  $Y^h$  (YY,  $YY^h$ , ou  $Y^hY^h$ ) é inviável, e os embriões abortam 25-50 dias após a polinização, sugerindo que os

tipos de cromossomos Y são semelhantes e que em ambos falta um gene essencial, que é funcional no cromossomo X (Yu et al., 2008).

Populações de mamão silvestres são dióicas, com metade das plantas masculina e metade feminina, enquanto o mamoeiro cultivado é predominantemente ginodioicos, com dois terços de plantas hermafroditas e um terço de plantas femininas, embora existam variedades dioicas (Colunga-García Marín e Zizumbo-Villarreal, 2004). A diversidade das formas florais observada no mamoeiro pode ser dividida, basicamente, em três tipos bem diferenciados: flor hermafrodita, que caracteriza as plantas hermafroditas; flor estaminada, que caracteriza as plantas masculinas; e flor pistilada, que caracteriza uma planta feminina típica (Storey, 1941). De acordo com Hofmeyr (1939) e Storey (1941), podem ocorrer alterações nos tipos florais, causadas, principalmente, por fatores genéticos e ambientais. A amplitude dessa transição de sexos é dependente da variedade, idade da planta e do ambiente (Storey, 1941).

As flores femininas são consideradas estáveis, unissexuais e pistilo pentacarpelado com ausência total do órgão masculino. As flores masculinas são unissexuais, com um pistilo rudimentar e ausência de estigmas e 5 pares de estames. Os estames da flor masculina são órgãos florais estáveis, ao contrário do pistiloide, que, por condições adversas de temperatura, se desenvolve podendo ocorrer, assim, autofecundação e, conseqüentemente, produção de frutos.

As flores hermafroditas são bissexuais, com duas séries de cinco estames cada uma, pistilo normalmente pentacarpelado. Em alguns casos, anteras e estigmas localizam-se numa mesma altura, o que pode permitir a ocorrência de autofecundação, porém as anteras podem se localizar abaixo do estigma, reduzindo as chances ocorrer este processo (Hofmeyr, 1939). A reversão sexual em plantas com flores hermafroditas ocorre nos estádios iniciais de desenvolvimento do carpelo em que há a diferenciação do órgão feminino, mas ele não se desenvolve. Um dos principais motivos para ocorrer esta anomalia é o aumento da temperatura. A flor hermafrodita passa a ter característica de flor masculina, com presença de um carpelo rudimentar (Storey, 1941).

### 3.5. Genes MADS-Box em plantas

Os genes MADS-box codificam proteínas com um domínio de ligação ao DNA altamente conservado, aproximadamente 180 nucleotídeos na região N-terminal, denominado MADS. O domínio MADS tem aproximadamente 60 aminoácidos e reconhece uma sequência específica [CC(AT)<sub>6</sub>GG] denominada CArG-box, na região regulatória de seus genes alvo (Hayes et al., 1988; Shore e Sharrocks, 1995; Khan et al., 2012).

Fatores de transcrição com domínio MADS estão envolvidos no controle de muitos processos do desenvolvimento de plantas com flores, processos que vão desde o desenvolvimento do saco embrionário e do grão de pólen até o desenvolvimento da raiz, flor e fruto (Gramzow et al., 2010; Becker e Theissen, 2003).

A acronímia MADS é derivada dos quatro membros fundadores desta família de genes: *MINICHROMOSOME MAINTENANCE 1* de *Saccharomyces cerevisiae*, *AGAMOUS* da *Arabidopsis thaliana*, *DEFICIENS* de *Antirrhinum majus* e *SERUM RESPONSE FACTOR* de *Homo sapiens* (Messenguy e Dubois, 2003; Schwarz-Sommer et al., 1990). A expansão e diversificação de genes MADS-box em plantas está intimamente ligada à evolução de novas estruturas, tais como flores, frutas e sementes (Theissen et al., 2000).

As análises filogenéticas mostraram que a duplicação do gene MADS-box ancestral ocorreu antes da divergência de plantas, animais e fungos, dando origem a duas linhagens, caracterizadas como tipo I e tipo II (Alvarez-Buylla et al., 2000). Dados sugerem que os MADS-box se originaram de uma sequência de DNA que codifica uma região da subunidade A da enzima topoisomerase IIA (Gramzow L. et al., 2010). Em plantas, como resultado de uma duplicação e posterior divergência, surgiram dois tipos de genes MADS-box, denominados tipo I e tipo II, sendo estes grupos distinguidos pela estrutura do domínio, taxas de evolução, função do desenvolvimento e do grau de redundância funcional.

Genes MADS-box do Tipo I, correspondente ao grupo de genes  $M\gamma$ ,  $M\beta$  e o  $M\alpha$ , apresentam um ou dois éxons e ausência do domínio K nas proteínas. Estudos permitiram demonstrar que os genes pertencentes a esse grupo têm funções relacionadas com a gametogênese feminina e com o desenvolvimento de sementes (Wuest et al., 2010).

Genes MADS-box do Tipo II, atuam principalmente no desenvolvimento floral, embriogênese e no desenvolvimento do fruto (Kaufmann et al., 2005). Este grupo é caracterizado por uma região conservada chamada de MIKC com quatro domínios característicos- MADS de ligação ao DNA (M), Intermediário (I), queratina-like (K) e carbox terminal (C)- distribuídos desde a região amino-terminal (N) até a região carboxi-terminal da proteína sendo responsáveis pela conformação e especificidade funcional das proteínas MADS-box (Henschel et al., 2002; Parenicova et al., 2003; Robles e Pelaz, 2005). Genes MADS-box do Tipo II apresentam em média sete éxons (De Bodt *et al.*, 2003).

Com base nas características estruturais desse domínio, os genes MADS-box do tipo MIKC foram divididos em MIKC<sup>c</sup> (forma canônica) e MIKC\*, que apresenta alterações no domínio K, possivelmente pela duplicação de éxons desta região, formando uma rede de interações de proteínas responsável pela maturação do pólen (Adamczyk e Fernandez, 2009). Genes com funções similares e com padrões de expressão similares em diferentes espécies estão geralmente relacionados e pertencem ao mesmo subtipo de genes MADS-box (Becker e Theissen, 2003).

Duplicações em grande escala podem favorecer a manutenção dos genes Tipo II, principalmente pelo fato de esses genes apresentarem subunidades proteicas como o domínio K, que facilita as interações proteína-proteína, o que pode não ocorrer se a duplicação for em pequena escala ou tandem (Freeling et al., 2009; Edger et al., 2009). Duplicação gênica é um processo crucial na diversificação e adaptação biológica. Genes duplicados representam adição de variabilidade sob a qual uma combinação de mutação e seleção natural pode levar ao surgimento de novas funções (Alvarez-Buylla et al., 2000; Lawton-Rauh, 2003)

Em um estudo com 17 espécies com o genoma sequenciado foram identificados 2.060 genes MADS-box, dos quais 1.057 foram classificados como tipo I e 990, como tipo II. *Solanum tuberosum* é a espécie que tem maior número de genes MADS-box, tem sido identificados um total de 265, dos quais 173 genes pertencem ao MADS tipo I e 86 são do tipo II.

Apesar de o conteúdo gênico do *Carica papaya* ser considerado pequeno pelo tamanho do seu genoma, quando comparado com *Arabidopsis*, o número de genes MADS identificados foi surpreendente, sendo que, dos 262 genes MADS-

box identificados, 229 são genes do tipo I e somente 30 são genes do tipo II (Gramzow et al., 2010; Ming et al., 2008).

### 3.6. Redes reguladoras de genes no desenvolvimento floral

Os processos relacionados ao desenvolvimento da planta são controlados por redes fortemente coordenadas de reguladores, conhecidos como redes reguladoras de genes (RRG), que ativam e reprimem a expressão do gene dentro de um contexto espacial e temporal (Jacob e Monod, 1961; Hart, 2002).

Em *Arabidopsis thaliana*, os principais componentes dos RRG que controlam os principais processos de reprodução das plantas, tais como a transição floral e a especificação da identidade órgão floral, foram identificados pela primeira vez em mutantes funcionais que perderam a função que afeta esses processos do desenvolvimento (Blázquez et al., 2006). Grande parte desses genes reguladores codifica fatores de transcrição (FTs), muitas vezes membros da família MADS-box (Pajoro, 2014).

A transição do desenvolvimento vegetativo para o reprodutivo em plantas é regulada por sinais endógenos e estímulos ambientais (Borner et al., 2000). Várias vias genéticas, incluindo fotoperíodo, via autônoma, vernalização e vias induzidas por giberelina, convergem em um pequeno conjunto de genes centrais reguladores da floração, que, em *Arabidopsis thaliana*, *FLOWERING LOCUS T* (*FLT*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*) e *LEAFY* (*LFY*), integram esses sinais das múltiplas vias para promover a transição da floração (Boss et al., 2004; Sung e Amasino, 2004)

*CONSTANS* (*CO*) e *FLOWERING LOCUS C* (*FLC*) agem antagonicamente regulando florescimento. *FLC* atua como repressor da floração e medeia as vias autônomas e de vernalização, enquanto *CO* é um ativador da floração e medeia a via do fotoperíodo. *CO* e *FLC* juntos regulam a expressão de genes da via de integradores da floração *FLOWERING LOCUS T* (*FLT*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), e *LEAFY* (*LFY*) (Borner et al., 2000; Samach et al., 2000). Esses genes integradores da floração podem ativar outra classe de genes, conhecidos como genes de identidade do meristema floral, tais como *APETALA 1* (*AP1*), *FRUITFUL* (*FUL*), *CAULIFLOWER* (*CAL*), *LYF* e *SEPALLATA 4* (*SEP4*). As atividades desses

genes de identidade floral promovem a transição dos meristemas vegetativos em meristemas reprodutivos pela regulação de uma terceira classe de gene, que são os genes de identidade dos órgãos florais. Esses genes controlam diretamente a identidade floral, promovendo o desenvolvimento sequencial dos primórdios florais em cada verticilo floral (Tan e Swain, 2006).

As múltiplas vias que controlam o tempo de floração podem ser divididas em duas rotas: a rota que promove a indução floral, que inclui sinais ambientais e endógenos, e a rota que possibilita a indução floral, que atua na regulação da habilidade do meristema em responder a distintos sinais como vernalização, estágio de desenvolvimento e via autônoma. As vias que promovem à indução floral irão ativar a expressão dos genes integradores da floração, e a via que possibilita a indução floral pode aumentar ou diminuir a eficácia dos repressores, e conseqüentemente, inibir ou promover o florescimento (Boss et al., 2004).

O crescimento vegetativo em *Arabidopsis* é mantido pela repressão da função dos genes do florescimento. O gene *SHORT VEGETATIVE PHASE* (SVP) é um repressor que interage com FLC para reprimir diretamente a expressão SOC1 (Li et al., 2008) e, com isso, regular o tempo do florescimento pela manutenção do crescimento vegetativo. O gene *TERMINAL FLOWER 1* (TFL1) é outro importante repressor que atrasa o florescimento e regula o desenvolvimento da planta pela manutenção do estágio indeterminado do meristema de inflorescência, inibindo a formação das flores (Shannon e Meeks-Wagner, 1991). Ao iniciar a floração, as interações entre os genes específicos do florescimento e genes requeridos para a manutenção do meristema, entre eles o gene *WUSCHEL*, garantem que as florações em *Arabidopsis* irão se desenvolver mantendo sua integridade estrutural e funcional nos quatro verticilos (Laux et al., 1996).

### **3.7. Bioinformática e Análises Filogenéticas no Melhoramento de Plantas**

Filogenia é o estudo da relação evolutiva entre grupos de organismos. A evolução é considerada um processo de ramificação e pode ser visualizada em uma árvore filogenética (Hennig, 1966). Cada ramo representa a persistência de uma linhagem genética através do tempo, e cada nó representa o nascimento de uma nova linhagem (Yang e Rannala, 2013). Uma maneira pela qual as plantas podem responder rapidamente a mudanças ambientais externas é modificando a expressão dos genes envolvidos no crescimento e em outros processos. Portanto,

compreender como isso acontece irá nos ajudar a melhorar a capacidade de sobrevivência de algumas culturas quando expostas à seca ou a outras condições ambientais extremas (Ahuja et al., 2010; Weston et al., 2008).

O estudo da morfologia floral no contexto evolutivo tem sido sustentado pela análise genética, que permite compreender os mecanismos moleculares envolvidos na formação dos órgãos na planta e na evolução destas estruturas (Richards et al., 2009). A organização de grande parte das proteínas é hierárquica sendo dispostas em grupos denominados famílias e superfamílias, e essa organização reflete um processo ancestral e contínuo de duplicação e divergência gênica (Thornton, 2002). Os genes responsáveis por controlar o desenvolvimento, na maioria das vezes, são os genes Homeóticos e pertencem a um número limitado de famílias multigênicas que codificam fatores transcricionais. Os genes homeóticos são altamente conservados e tendem a compartilhar funções similares na morfogênese das flores. A grande maioria dos genes homeóticos florais pertence à família gênica MADS-box (Theissen et al., 2000).

Com as quantidades cada vez maiores de dados biológicos, a integração com a tecnologia da informação tornou-se essencial. A contribuição da bioinformática para o desenvolvimento genômico funcional de plantas é fundamental. A identificação e a determinação da posição filogenética dos genes MADS-box em espécies vegetais permitem a formulação de hipóteses sobre o desenvolvimento morfológico destas espécies e a função de seus respectivos genes. O conhecimento do conjunto completo de genes MADS-box ao longo da planta permitirá uma compreensão mais profunda da evolução dos principais grupos de plantas da terra, e principalmente das estruturas que os define (Gramzow e Theissen, 2013).

Tecnologias como a genômica, proteômica e a bioinformática vêm contribuindo para a aceleração do conhecimento científico. Tecnologias de sequenciamento de DNA tornaram-se técnicas essenciais para diversas áreas da ciência, trazendo benefícios para as áreas de conhecimento que abrangem arqueologia, antropologia, genética, microbiologia, biotecnologia, biologia molecular e ciências forenses, entre outras .

As tecnologias de sequenciamento de nova geração, conhecidas como plataformas NGS, revolucionaram a biologia moderna ao permitir o sequenciamento de genomas completos de forma mais rápida, por meio do

sequenciamento em paralelo, e a custos mais baixos (Glenn, 2011). O sequenciamento por Sanger continua sendo a tecnologia que gera dados com maior qualidade, sendo ainda muito utilizado para detectar pequenas mutações nos genomas, os SNPs. Nos estudos de taxonomia e identificação de espécies, essa metodologia é extensamente utilizado através do sequenciamento da região ITS do DNA ribossomal (Chen et al., 2001).

Muitos projetos de sequenciamento genômico têm sido realizados com o objetivo de compreender melhor os sistemas biológicos. Genomas completos de plantas já foram totalmente sequenciados, como a *A. thaliana* (*A. thaliana Genome Initiative*, 2000), *Carica papaya* (Ming et al., 2008), *Coffea canephora* (Denoëud et al., 2014), *Humulus lupulus* (Natsume et al., 2015), *Glycine max* (Schmutz et al., 2010), *Theobroma cacao* (Argout et al., 2011), *Populus trichocarpa* (Tuskan et al., 2006), *Oryza sativa* (*International Rice Sequencing Project*, 2005), Melancia (Shaogui et al., 2013), Dendezeiro (Singh et al., 2013) gerando uma grande quantidade de informações.

## 4. MATERIAL E MÉTODOS

### 4.1. Material Genético

O material vegetal utilizado no experimento foram folhas de plantas já sexadas, hermafroditas, de quatro genótipos de mamoeiro - *Sunrise Solo 72/12*, *JS12*, *Golden* e híbrido UENF-Caliman 01 - e femininas do genótipo Cariflora. As variedades estavam plantadas na área de plantio comercial na fazenda Caliman Agrícola S/A, localizada no município de Sooretama/ES (Latitude: 19° 11' 49" S; Longitude: 40° 05' 52" O).

A variedade 'Sunrise Solo 72/12' é procedente do Havaí, EUA, e, por meio de seleção massal, foi introduzida em 1986 nas fazendas Experimentais de Linhares e Viana, pela antiga Empresa Capixaba de Pesquisa Agropecuária (EMCAPA). Essa variedade é conhecida comumente como mamão Havaí, amplamente disseminada nas regiões produtoras do Espírito Santa.

A variedade 'Golden' é uma variedade brasileira oriunda de uma mutação da linhagem "Sunrise Solo", selecionada pela Caliman Agrícola S/A no município de Linhares-ES. Apresenta frutos e plantas de coloração verde-clara.

A linhagem 'JS12' foi selecionada pela Empresa Baiana de Desenvolvimento Agrícola para a região do Recôncavo Baiano. Em 1998, foi introduzida no banco de germoplasma da UENF, por meio de sementes provenientes da Embrapa Centro Nacional de Pesquisa da Mandioca e Fruticultura, de Cruz das Almas, BA.

'UENF-Caliman 01' é o primeiro híbrido desenvolvido no Brasil pela UENF e Caliman Agrícola S/A, no município de Linhares – ES, para as regiões norte e noroeste fluminense e norte litorâneo do Estado do Espírito Santo. O híbrido é resultante do cruzamento de 'Sunrise Solo 72/12' com o 'JS12'.

'Cariflora' é uma variedade dióica tolerante ao *Papaya ringspot vírus* (PRSV). Resultados obtidos por Conover & Litz (1979) sugerem que o caráter tolerância tem herança poligênica. Estes autores consideram possível o desenvolvimento de cultivares tolerantes ao vírus, com resistência horizontal, como uma alternativa para reduzir as perdas no cultivo.

#### **4.2 Extração do DNA genômico**

Foram coletadas cinco folhas de cada genótipo de mamoeiro e armazenadas no -80°C. Um total de 150 mg de amostra macerada foi utilizada para a extração DNA total, utilizando o kit *Qiagen*, seguindo recomendações do fabricante. Para a quantificação do material extraído, utilizou-se o *NanoDrop ND-100 (Thermo-Fisher Scientific)* com leituras de absorvância na faixa de 230, 260 e 280nm. Para análise da qualidade, foi aplicado 1µL de DNA genômico em gel de agarose (1,5%), corado com *Gel Red®* e azul de bromofenol. Após as análises preliminares de qualidade, o material extraído foi armazenado no freezer a -20°C.

#### **4.3 Seleção de genes homeóticos e desenvolvimento de primers**

Em uma análise previa, foram identificadas 76 sequências de aminoácidos com o domínio MADS-box no genoma do mamoeiro. A classificação dessas 76 sequências foi feita por sucessivos alinhamentos entre aminoácidos de *Arabidopsis* e *Carica papaya* no programa de CLUSTALW, com parâmetros prédefinidos (Figueiredo, 2012).

Os iniciadores para análise de sequenciamento foram construídos utilizando como ferramenta principal o software PRIMER3 plus (Rozen e Skaletsky, 2000) e para a avaliação os programas NetPrimer (Premier Biosoft Internacional) e OligoExplorer 1.4 Beta (Gene Link). Os seguintes parâmetros físico-químicos desejáveis nos iniciadores foram respeitados: ausência de formação de "hairpin" e dímeros, baixa complementaridade interna, estrutura secundária ausente ou restrita, baixa estabilidade da extremidade.

Os iniciadores foram desenhados com uma distância média aproximada de 150pb, do início da sequência gênica e do final para assegurar o sequenciamento completo da sequência codificante.

#### **4.4 Reação em cadeia da polimerase - PCR**

O DNA extraído das amostras de folha do mamoeiro foi utilizado como molde na reação em cadeia da polimerase para a amplificação das regiões determinantes dos possíveis genes MADS-box do mamoeiro.

Os produtos de PCR foram submetidos à eletroforese em gel *MetaPhor* 1% (m/v), sob corrente elétrica de 100v, em tampão TAE (1mm EDTA pH 8,0; 40mm Tris pH8,0; 20mm de ácido acético), por 60 minutos, utilizando como referência a escala 1kb DNA *ladder* (Kasvi). As amostras foram coradas com *Gel Red®* (*Uniscience*) e visualizados sob luz UV e as imagens capturadas, pelo fotodocumentador.

Amostras de 60ng do produto de PCR com seus respectivos *primers* foram enviadas para a ACTGene Análises Moleculares de sequenciamento utilizando o equipamento *ABI-Prism 3500 Genetic Analyzer* (*Applied Biosystems*). Foram enviadas amostras em triplicata de ambas, juntamente com os respectivos primers.

#### **4.5 Análises das sequências**

As regiões iniciais e terminais das sequências diretas e reversas foram editadas em ambas as fitas em triplicata com a ajuda do *BioEdit Sequence Alignment Editor* 7.1.9.

As sequências foram analisadas no programa *CodonCode Aligner*, versão 6.0.2, para a obtenção dos *contigs*.

#### **4.6 Análises Filogenéticas moleculares**

Para o alinhamento global dos *contigs* e dos aminoácidos, utilizou-se o programa Clustal X 2.1(Thompson et al.,1997), com iteração em cada etapa de alinhamento. As árvores filogenéticas foram geradas pelo programa MEGA *Beta* 6 (Molecular Evolutionary Genetics Analysis) (Tamura et al., 2007) e o CLC 7.5 (*Genomics Workbench Portal*). As árvores filogenéticas para as sequências de nucleotídeos foram elaboradas pelos métodos de *Neighbor-Joining* (NJ) e, para

dar suporte aos clados obtidos, empregou-se o teste de *bootstrap* (Felsenstein, 1985), utilizando-se 10.000 réplicas. Para os métodos NJ, utilizou-se o modelo evolutivo com base no método JTT (Jones et al., 1992).

## 5. RESULTADOS E DISCUSSÃO

### 5.1 Coleta do material vegetal e extração do DNA genômico

A extração de DNA é o primeiro passo na utilização de técnicas moleculares. Este processo é parte fundamental para se obter alta eficiência de amplificação nos protocolos que usam a Reação em Cadeia da Polimerase (PCR). A integridade do DNA genômico extraído foi conferida por eletroforese em gel de agarose 1%, - Figura 3 – tendo sido constatado que não houve contaminação por RNA.

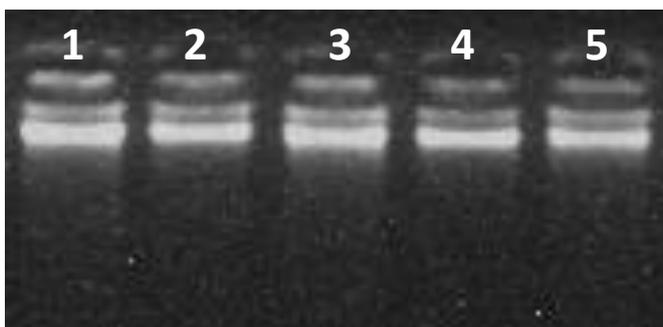


Figura 3: Extração do DNA genômico dos cinco genótipos do banco de germoplasma de mamoeiro, em gel de agarose 1%. 1: UENF Caliman 01, 2: JS12, 3: Sunrise Solo, 4: Golden e 5 : Cariflora

### 5.2 Iniciadores utilizados nas análises

Em razão da importância no desenvolvimento e identidade floral, os possíveis genes MADS de mamoeiro, *FUL/Agamous like 8* e *SOC1/Agamous like 20*, foram selecionados para o desenvolvimento desse trabalho. As sequências de

nucleotídeos, referentes aos genes MADS-box que foram retiradas do *Phytozome*, estão descritas na Figura 4 e, marcado em azul, está o CDS (sequência codificadora) da sequência genômica.

*Phytozome* é uma plataforma genômica comparativa de plantas do Departamento de Energia dos Estados Unidos e do Instituto de Pesquisa *Joint Genome*. Atualmente, o *Phytozome v.11* fornece acesso a 58 genomas de plantas que foram sequenciadas e anotadas, dos quais 52 foram agrupados em famílias gênicas. Cada gene tem sido anotado utilizando plataformas como *PFAM*, *KOG*, *KEEG*, *PANTHER* e *GO*. O genoma do *Carica papaya* está depositado no *Phytozome* e tem aproximadamente 135 Mb dispostos em 4114 contigs. Em relação aos genes MADS-box, as sequências depositadas não têm o domínio *Keratin-like* (K) responsável por interações proteína-proteína bem definidas. A sequência depositada no banco de dados *Phytozome* para ambos os genes apresenta somente um único éxon referente ao domínio MADS (M) altamente conservado SRF.

A tabela 1 mostra as sequências dos iniciadores dos genes MADS-BOX utilizados para a amplificação

<b>Gene SOC1</b>	<b>Gene FUL</b>
GGTGTATGTGTAGGCACTGTTTAGTTGCTCAA AATTTTCAGGTTCTTTTGTGTTGATTTCTTTCT TAAATTTCTTTCGCCTTCATTTTTTACTCTCT TGAGCAATAAAAACAAGATCTGGGGGGTTTTT GCATCTGACTTGGTTGACTTTCTTTAGTACTA TCCTTAATCTCTTACTTTGCACACCTTCTCTT TGTTGATAAGAAAATCAAATATGCATTGAAGA GCTTCTTAATTTGCTTTCTTTAGTGTGTTGTG TGTTAAATTTTAGTCTTTGTTTCACTCTATGA CATAAAGTAAGTTGGTGTATATATATATATA TATATATAGATGAGGTTGAATATTTCTTTTG TTGTTCAAAAGGTTGTAAAGATGGTGAGAGGG AAGACACAAATGAAGCGGATAGAGAATGCTAC AAGCCGGCAAGTGACCTTCTCCAAGCGCAGAA ATGGTCTGCTGAAGAAAGCATTTGAGCTTTCA GTTCTTTGTGATGCTGAGGTTGCCCTTATCAT CTTTTCTCCCAGAGGAAAGCTCCATGAATTTG CGAGTGCAAGGCTCATAACTCTTTAACCTTC TATCTTTTCTGTTTCATTAATTGATTATCTCT TATTAATTCAATCACTATATTCTTTTGCAATA TTCATTGTACATATATGTATACGTATATGATC TGACATTACTCACTTTTATTTTAACACACGCA	TTTTGCTTCTCAAACAGCAAAAACCAGTACGA AAGTAGAAACCTTTCTATATTTGCCATATCCA TCACTATATTCCCCACTCTTTCATCAATTCTT CTTTCTTCTCTCCTGGGTCTTGTTCTTTTAT GCCAAAAAAGATGGCGAAACTTTAGGGTTCTT TCAATCCTTGTTTGCCTGGTGAGAAAATTAGA AGGAACTAAGAGAGGGGAAGGAGAGAGAGAG AGAGGTGGGTAAATGGGGAGAGGAAGGGTTCA GTTAAAACGGATCGAGAACAAGATCAGCAGAC AAGTAACTTTTTCCAAGAGGAGAACCAGTTTA CTCAAGAAGGCTCATGAGATCTCCGTGCTGTG CGAAGCTGATGTTGCTCTGATTGTCTTCTCCA CCAAAGGAAAGCTCTATGAGTACTCTACTGAT TCAAGGTTTGTCTTTTTTTCTTCTCCTTTCA TACATAAATTCTTTACCTTTACCTGGTATCTG CTTATTCTTTAACTGTTTTTAACCTTATTTGCT GCTTTTCTACCTCTGGGTTTTTTGTTATAATTA TGTTTTTTTTCGTGGGTCTTCAAACATACTTG GTTCTTTTTTCTTTTTCCGTACAATGTCATGG AAAAAC

Figura 4: Sequências genômicas dos possíveis genes *SOC1* e *FUL* retiradas do banco de dados *Phytozome* (<https://phytozome.jgi.doe.gov/pz/portal.html>). Em azul está o CDS da sequência.

**Tabela 1:** Iniciadores dos genes MADS box de mamoeiro, construídos pelo *PRIMER 3 plus*, referentes aos possíveis gene *SOC1* e ao gene *FUL* de mamoeiro.

Contig		Sequência do primer (5` - 3`)	Tamanho do fragmento (pb)	Gene
evm.TU.supercontig_18.166	R	CTTTGCACACCTTCTCTTTGTTG	580	Soc1
evm.TU.supercontig_18.166	F	CATGGAAACCCAGGAGCTACG		
evm.TU.supercontig_14.223	R	GACAGCACATGCCCCAGTT	840	FUL
evm.TU.supercontig_14.223	F	GCTTTTCGGTCAGCTGCAG		

### 5.3 Reações em cadeia da Polimerase PCR

É possível observar que os fragmentos de DNA das amostras foram amplificados e exibiram especificidade satisfatória. Os iniciadores específicos para o gene *SOC1* amplificaram um produto de aproximadamente 580pb, e os iniciadores específicos para o gene *FUL* amplificaram um produto de aproximadamente 840pb. O tamanho do fragmento gerado após a reação de PCR está condizente com o tamanho do fragmento descrito na construção pelo programa *primer 3 plus* (Figura 5).

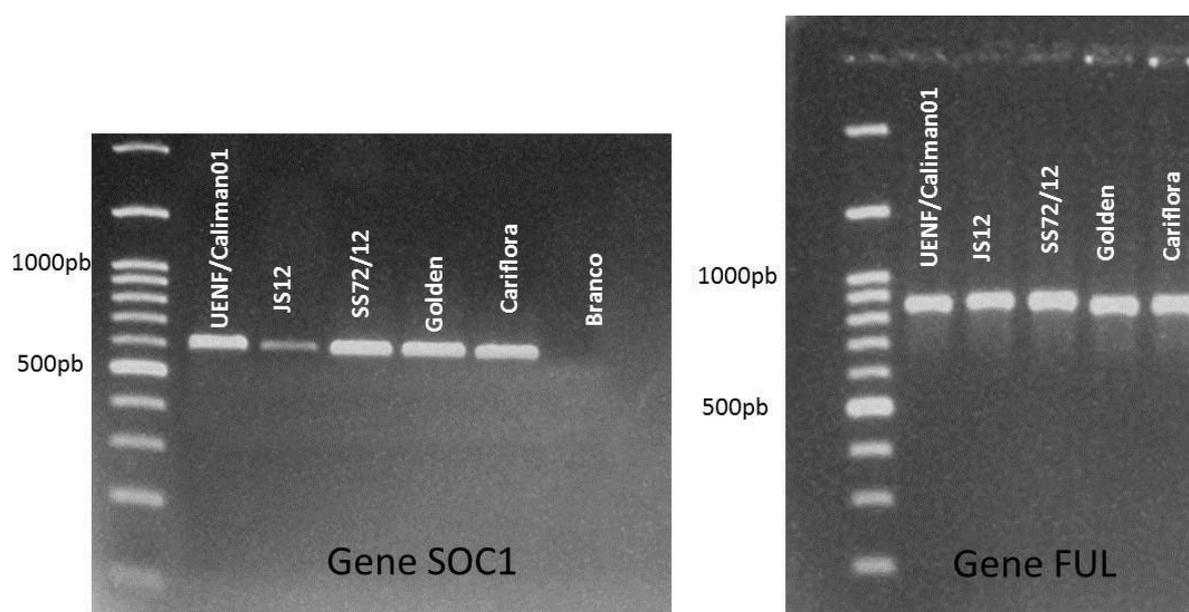


Figura 5: Produtos de amplificação dos genes *SOC-1* (Gene 2) e do *FUL* (Gene 4) em gel *Methaphor* 1%. Foram aplicados 1  $\mu$ L do produto e 2  $\mu$ L do *Ladder* 1000pb KASVI (100-3000pb).

A quantidade de material de partida (DNA molde) para a reação de sequenciamento é extremamente crítica para a obtenção de boas sequências. As quantidades variam de acordo com o tamanho dos fragmentos que se pretende sequenciar.

Após a amplificação dos fragmentos específicos, as amostras foram quantificadas no *nanodrop* e diluídas na concentração 30 -60ng, estipulada pela empresa ACTGene Análises Moleculares.

#### **5.4 Sequenciamento e análises de bioinformática**

Destaca-se que ao longo da primeira etapa de trimagem dos dados o mais importante é identificar os iniciadores para localizar e delimitar o início e o fim da região gênica.

Identificando a região dos primers e retirando as sequências adjacentes ao fragmento amplificado, iniciou-se a formação dos contigs para posterior montagem. Com o auxílio do CodonCode, formou-se um contig com as três sequências *Forward* e um contig com as três sequências *Reverse* para cada genótipo analisado. Ambos os *contigs*, *Forward* e *Reverse*, de cada gene foram alinhados e montados, formando assim um único *contig* de sequências nucleotídicas com o tamanho de 580pb para o possível gene SOC1 e com 840 pb para o possível gene FUL em cada genótipo. No final das análises de trimagem e montagem, formou-se 10 contigs.

Os domínios MADS -box de todos os dez *contigs* foram identificados com o auxílio do banco de dados biológicos on-line *Pfam*. A sequência de aminoácidos dos CDS foi separada e comparada com o banco de dados *Uniprot* e *Swiss-Prot*. Sequências com identidade maior que 70% e um *e-value* variando de -40 a -29 foram separadas para a análise filogenética. A montagem em nível de nucleotídeos pode ser chamada de anotação estrutural (Tabela 2). A anotação em nível de proteína pode ser denominada de anotação funcional, uma vez que ela associa uma função biológica à região anotada.

Tabela 2: Visualização da anotação estrutural sequência nucleotídica e da sequência de aminoácidos do domínio MADS-Box *SERUM RESPONSE FACTOR* SRF para cada genótipo do possível gene *FRUITFUL* e *SUPRESSOR OF OVEREXPRESSION OF CONSTANS1*.

Gene /genótipo	Sequencia Nucleotidica	Sequencia aminoácido	Domínio MADS
FULL/Calimosa	ATGGGGAGAGAAAAGGGTTCAGAGAAAA CGGAGCGAGAAACAAGATCACCAGACAA GAGACTTTTTCCCAGAGAAGAACCAGTT TACACTCGAGGGGCGCACGAGATCTCCG CGCGGTGCGAAGAGGATGTGGCTCTGA TTGTGTTCTCCCCCAAGGAAAGCTCTA TGAGAACTCTCCTGAGTCACGGGGTGT GCTTTTTTCTTCTCCTTTACACACAA	MGRERGRQRKRS ENKITRQETFSQR RTSLHSRAHEISA RCEEDVALIVFSP QGKLYENSPESR GVLFFFSFHTQ	SRF
FULL/Cariflora	ATGGGGAGAGGAAGGGTTCAGTTAAAA CGGATCGAGAAACAAGATCAGCAGACAA GTAACTTTTTCCAAGAGGAGAACCAGTT TACTCAAGAAGGCTCATGAGATCTCCGT GCTGTGCGAAGCTGATGTTGCTCTGATT GTCTTCTCCACCAAAGGAAAGCTCTATG AGTACTCTACTGATTCAAGGTTTGTCT TTTTTCTTCTCCTTTACATACATAA	MGRGRVQLKRIE NKISRQVTFKRR TSLKKAHEISVL CEADVALIVFSTK GKLYEYSTD SRF VLFFFSFHT*	SRF
FULL/SS72/12	ATGGGGAGAGGAAGGGTTCAGTTAAAA CGGATCGAGAAACAAGATCAGCAGACAA GTCACTTTTTCCAAGAGGAGAACCAGTT TACTCAAGAAGGCTCATGAGATCTCCGT GCTGTGCGGACGGCTACGTAGCACTGAT TGCTTCTCCACCAAAGGAAAGCTCTAT GAGTACTCTACTGATCCAAGGTTTGTTC TTTTTCTTCTCCTTTGCATACATAA	MGRGRVQLKRIE NKISRQVTFKRR TSLKKAHEISVLS DGYVALIVFSTKG KLYEYSTDPRFVL FFFSLHT*	SRF
FULL/JS12	ATGGGGAGAGGAAGGGTTCAGTTAAAA CGGATCGAGAAACAAGATCAGCAGACAA GTAACTTTTTCCAAGAGGAGAACCAGTT TACTCAAGAAGGCTCATGAGATCTCCGT GCTGTGCGAAGGCTATGTTGCTCTGATT GTCTTCTCCACCAAAGGAAAGCTCTATG AGTACTCTACTGATCCAAGGTTTGTCT TTTTTCTTCTCCTTTACATACAAAA	MGRGRVQLKRIE NKISRQVTFKRR TSLKKAHEISVL CEGYVALIVFSTK GKLYEYSTDPRF VLFFFSFHTK	SRF
Golden/Full	ATGGGGAGAGGAAGGGTTCAGTTAAAA CGGATCGAGAAACAAGATCAGCAGACAA GTAACTTTTTCCAAGAGGAGAACCAGTT TACTCAAGAAGGCTCATGAGATCTCCGT GCTGTGCGAAGGTGAGGTTGCTCTGAT TGCTTCTCCACCAAAGGAAAGTTCTAT GAGTCCCCTACTGTTTCAGGGTGGTC CTTTTTTCTTCTCCTTTACATAATAA	MGRGRVQLKRIE NKISRQVTFKRR TSLKKAHEISVL CEGEVALIVFSTK GKPFYESPTV SGL VLFFFSFHNK	SRF
SOC1/Calimosa	ATGGGGAGAGGGGAAGAGACAGATAAAG CGGATAGAGAACGCGATCTGCAGGCAA GTGACTTTCTCCAAGAGAAGAACACGT GTTACACGAAAAGCTCACGAGATCTCT GTGCTGTGTGATGCTGAGATTGTCGCT CATAGTCTTTTCTCCAGAGGAAAGCTC TCTATGAACCTCTCCAGTGAAGTGGACTG TGCTCTTTTTCTTCTCCTTTCTCACAT	MGRGKRQIKRIEN AICRQVTFKRR RVHTKAHEISVLC DAEIVAHLSFSQR KALYELSSDSDCA LFLLPFSH	SRF
SOC1/SS72/12	ATGGGTGAGAGGGGAAGACACAAATGAAG CGGATAGAGAATGCTACAAGCCGGCAA GTGACCTTCTCCAAGCGCAGAAATGGT CTGCTGAAGAAAGCATTGAGCTTTCAG TTCTTTGTGATGCTGAGGTTGCCCTTAT CATCTTTTCTCCAGAGGAAAGCTCCAT GAATTTGCGAGTGCATGGTCTCATATCT CATTACCCATCTTCTTGTCTGTGCATA GAGTGAA	MVRGKTQMKRIE NATSRQVTFKRR RNGLLKAFELSV LCDAEVALIIFSPR GKLEHAFASAWSHI SFTHLPLCLS*SE	SRF
SOC1/JS12	ATGGGTGAGAGGGGAAGACACAAATGAAG CGGATAGAGACTGCTACAAGCCGGCAA GTGACCTTCTCCAAGCGCAGAAATGGT CTGCTGAAGAAAGCATTGAGCTTTCAG TTCTTTGTGATGCTGAGGTTGCCCTTAT CATCTTTTCTCCAGAGGAAAGCTCCAT GAATCTGCGAGTGCAGGTTCTCATAACT CATTACCCATCCTTCTTGTCTGTGCATA GAGTGAA	MVRGKTQMKRIE TATSRQVTFKRR RNGLLKAFELSV LCDAEVALIIFSPR GKLEHESASARSH NSFTILPLCLS*SE	SRF
SOC1/Cariflora	ATGGGTGAGAGGGGAAGACACAACTGAAG CGGATAGAGACTGCTACAAGCCGGCAA GTGACCTTCTCAGGCGCAGAAATGGT CCGCTGAAGAAAGCATCTGAGCTTTCAG GTTCTTTGTGATGCCGTGGTTGCCCTCA TCATCTTTTCTCCAGAGGAAACACCA AGAAATTGCGAGTGCAGGTTCTCATATC TCTATACACCTAACATCCTTGTCTGTCA TAGAGTGA	MVRGKTQLKRIET ATSRQVTFKRR NGPLKKAELSVL CDAVVALIIFSPR GKHQEIASARSHI SIHLTSLSVIE*	SRF
SOC-1/Golden	ATGGGTGAGAGGGGAAGACACAAATGAAG CGGATAGAGAATGCTACAAGCCGGCAA GTGACCTTCTCCAAGCGCAGAAATGGT CTGCTGAAGAAAGCATCTGAGCTTTCAG TTCTTTGTGATGCTGAGGTTGCCCTTAT CATCTTTTCTCCAGAGGAAAGCTACC ATGAATTTGCGAGTGCAGGTTCTCATAT ACTACTATACACCCAAACATCCTTGTACT GTCACATA	MVRGKTQMKRIE NATSRQVTFKRR RNGLLKKAELSVL LCDAEVALIIFSPR GKATMNLRVQGLI YYYYTPNILVLSL	SRF

*Serum Response Factor* (SRF) é um membro da superfamília de fatores transcricionais MADS-box (MCM1, AGAMOUS, DEFICIENS e SRF). Esta proteína regula a atividade de muitos genes precoces imediatos e participa na regulação do ciclo celular, apoptose, crescimento de células e diferenciação celular. De modo geral, o domínio MADS é responsável pela ligação ao DNA, ligam tanto como homo ou heterodímeros, enquanto o restante da proteína está envolvido em interações proteína-proteína. Tais interações podem ocorrer entre duas proteínas MADS-box ou entre uma proteína MADS-box e outra proteína não relacionada.

As sequências peptídicas da região altamente conservada do domínio MADS-box SRF de todos os genótipos de mamoeiro (JS12, SS72/12, Cariflora, Golden e Calimosa) para o gene *FRUITFUL* e para o gene *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* foram comparadas com o banco de dados *Uniprot*. A identificação de sequências homólogas fornece uma base para a análise filogenética e para reconhecimento de padrões nas sequências. As espécies que apresentaram similaridade estatisticamente significativa foram separadas e algumas delas foram utilizadas nas análises de inferência filogenética. As proteínas foram classificadas em grupos de acordo com as similaridades entre as sequências das espécies (Tabela 3 e 4). A análise do levantamento de sequências homólogas dos genes MADS-box FUL e SOC1 para os genótipos em análise revelou dados interessantes, destacando-se a similaridade com sequências AGL19 e a possível duplicação do gene FUL no genoma de *Carica papaya*.

Os dados referentes às dez sequências gênicas estruturais montadas nesse trabalho, referentes aos genótipos JS12, SS72/12, Golden, Cariflora e Calimosa, reafirmam a similaridade funcional quando comparadas com o banco de dados.

Na natureza, mutações randômicas e recombinação levam a diversidades genéticas. Durante a evolução de uma família de proteína, um gene ancestral pode ser duplicado uma ou mais vezes, com suas cópias divergindo e formando genes ditos parálogos (Stein, 2001).

A importância da duplicação de genes em fornecer material genético para a evolução biológica é reconhecida desde os anos de 1930. Genes duplicados são muitas vezes referidos como parálogos, os quais formam as famílias de genes. Posteriormente, muitos genes duplicados tornam-se pseudogenes, ou

seja, sequências de DNA derivadas de genes funcionais, que, por sofrerem determinadas mutações, se tornam não funcionais.

Tabela 3: Organismos que apresentaram similaridade estatisticamente significativa com o gene MADS-box FUL encontrados nos genótipos de mamoeiros quando comparados com o banco de dados *Uniprot*. As sequências de aminoácidos foram utilizadas inferência filogenética.

Query	Organismos	Proteínas	Sequências peptídicas	E-value	Score	Ident.
Q38876	<i>Arabidopsis thaliana</i>	Agamous-like MADS-box protein AGL8	MGRGRVQLKRIENKINRQVTFSKR RSGLLKKAHEISVLCDAEVALIVFS SKGKLFYEYSTD	7.7e-30	273	86.7%
Q39371	<i>Brassica oleracea</i>	Floral homeotic protein APETALA 1	MGRGRVQLKRIENKINRQVTFSKR RAGLFKKAHEISVLCDAEVALVIFS HKGKLFYEYSTD	1.7e-29	271	83.6%
Q41274	<i>Sinapis alba</i>	Agamous-like MADS-box protein AGL8	MGRGRVQLKRIENKINRQVTFSKR RSGLLKKAHEISVLCDAEVALVIFS SKGKLFYEYSTD	2.2e-29	270	83.3%
D1MDP9	<i>Vitis vinifera</i> (Grape)	Fruitful	MGRGRVQLKRIENKISRQVTFSKR RSGLLKKAHEISVLCDAEVALIVFS TKGKLFYEYSTD	3.5e-31	282	93.3%
W9RHT1	<i>Morus notabilis</i>	Agamous-like MADS-box protein AGL8-like protein	MGRGRVQLKRIENKISRQVTFSKR RSGLLKKAHEISVLCDAHVGLIIFST KKGKLFYEYSSD	2.2e-33	282	83.9%
A0A0B2PMH0	<i>Glycine soja</i>	Agamous-like MADS-box protein AGL8	MGRGRVQLKRIENKINRQVTFSKR RSGLLKKAHEISVLCDAEVALIVFS TKGKLFYEYSSD	3e-33	281	86.9%
A8D7L0	<i>Carica papaya</i> (Papaya)	FRUITFUL-like protein	MGRGRVQLKRIENKINRQVTFSKR RSGLLKKAHEISVLCDAEVALIVFS TKGKLFYEYSTD	1.9e-30	277	88.3%
B9N6N7	<i>Populus trichocarpa</i>	Apeta1 family protein	MGRGRVQLKRIENKISRQVTFSKR RTGLLKAHEISVLCDAVAVIVFS TKGKLFYEYSTD	2.1e-30	277	88.3%
G0LEW7	<i>Aethionema carneum</i>	FRUITFULL-like protein	MGRGRVQLKRIENKINRQVTFSKR RSGLLKKAHEISVLCDAEVALIVFS SKGKLFYEYSTD	6.2e-17	189	62.3%
A0A0B2R614	<i>Glycine soja</i> (Wild soybean)	Agamous-like MADS-box protein AGL8	MGRGRVQLKQIENKISRQVTFSKR RTGLRKKANEISVLCDAQVALIVFN AKGKLFYEYSSSRFVI	1.8e-18	190	60.0%
B9HIG9	<i>Populus trichocarpa</i>	AP1-like family protein	MGRGRVQLKRIENKINRQVTFSKR RTGLLKAHEISVLCDAEVALIVFS HKGKLFYEYS	1.5e-16	187	65.5%
Q0Q5E9	<i>Nicotiana tabacum</i>	Fruitfull-like MADS-box protein	MGRGRVQLKRIENKINRQVTFSKR RSGLLKKAHEISVLCDAEVALIVFS TKGKLFYEYATDS	2.7e-28	263	82.0%
D1MDP8	<i>Vitis vinifera</i> (Grape)	Apeta1	MGRGRVQLKRIENKINRQVTFSKR RTGLLKAHEISVLCDAEVALIVFST KKGKLFYEYSTD	6.2e-32	287	90.0%
A0A0B2QI18	<i>Glycine soja</i> (Wild soybean)	Agamous-like MADS-box protein AGL8 like	MGRGRVQLKRIENKINRQVTFSKR RSGLLKKAHEISVLCDAEVALIVFS TKGKLFYEYSSDP	6.4e-32	287	86.9%
X5DS60	<i>Coffea arabica</i>	FUL	MGRGRVQLKRIENKINRQVTFSKR RSGLLKKAHEISVLCDAEVALIVFS TKGKLFYEYATD	1e-30	279	86.7%
W9S545	<i>Morus notabilis</i>	Agamous-like MADS-box protein AGL8	MGRGRVQLRRIENKINRQVTFSKR RSGLLKKAHEISVLCDAEVALIVFS TKGKLFYEYSTNS	1.5e-34	289	86.9%

Tabela 4: Organismos que apresentaram similaridade estatisticamente significativa com o gene MADS-box SOC1 encontrados nos genótipos de mamoeiros quando comparados com o banco de dados *Uniprot*. As sequências de aminoácidos foram utilizadas na inferência filogenética.

Query	Organismo	Proteína	Sequências peptídicas	E-value	Score	Ident.
O82743	<i>Arabidopsis thaliana</i>	Agamous-like MADS-box protein AGL19	MVRGKTEMKRIENATSRQVTFKRRN GLLKKAFELSVLC-DAEVALVIFSPRSK LYEFSSSS	8.9e-17	180	62.9%
A9YTS4	<i>Sinningia speciosa</i>	SOC1-like protein 1	MVRGKVQMKRIENATSRQVTFKRRN GLLKKAFELSVLC-DAEVALVIFSQKGR LYEFSS	2.7e-17	183	65.0%
A0A0B0Q1E7	<i>Gossypium arboreum</i> (Tree cotton)	MADS-box SOC1-like protein	MVRGKTQMKRIENAASRQVTFKRRN GLLKKAFELSVLC-DAEVALIIFSTRGKL YEFSS	7.2e-17	175	63.3%
F5HPS4	<i>Gypsophila paniculata</i>	Suppressor of overexpression of constans 1	MVRGKTQMKRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVALIIFSPRGKLY EFASS	4.2e-36	314	96.7%
064665	<i>Arabidopsis thaliana</i>	MADS-box protein SOC1	MVRGKTQMKRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVSLLIIFSPKGKLY EFASS	6.5e-35	306	93.3%
Q52Z19	<i>Pisum sativum</i> (Garden pea)	Suppressor of Overexpression of Constans1	MVRGKTQMKRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVALIIFSPRGKLY EFSSS	3.4e-33	295	91.7%
G7L1F6	<i>Medicago truncatula</i>	MADS-box transcription factor	MVRGKTQMKRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVALIIFSPRGKLY EFSSSCS	2.3e-33	296	90.3%
A0A0B2P5Q7	<i>Glycine soja</i> (Wild soybean)	MADS-box protein SOC1	MVRGKTQMKRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVALIIFSPRGKLY EFASS SMQESI	6e-27	254	77.3%
O64645	<i>Arabidopsis thaliana</i>	MADS-box protein SOC1	MVRGKTQMKRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVSLLIIFSPKGKLY EFASS	3.6e-25	242	78.3%
Q9XJ60	<i>Oryza sativa</i> subsp. japonica (Rice)	MADS-box transcription factor 50	MVRGKTQMKRIENPTSRQVTFKRRN GLLKKAFELSVLCDAEVALVIFSPRGKLY YEFASASTQKTI	2.5e-26	250	74.2%
C9EF57	<i>Magnolia virginiana</i>	Suppressor of Overexpression of Constans1	MVRGKTQMRRIENATSRQVTFKRRN GLFKKAFELSVLCDAEVALVIFSPRGKLY YEFASSMHKTI	3.3e-26	249	72.7%
A2Q6H2	<i>Medicago truncatula</i>	Transcription factor, MADS-box	MVRGKTQMKRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVALIIFSPRGKLY EFSSS	1.4e-27	246	80.0%
U5GQE3	<i>Populus trichocarpa</i>	Uncharacterized protein	MVRGKTQMRRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVAVIVFSPRGKLY YEFSS R-HIEL	1e-26	246	72.7%
W9S6L4	<i>Morus notabilis</i>	MADS-box protein SOC1	MVRGKTQMKRIENAASRQVTFKRRN GLLKKAFELSVLCDAEVALVIFSPRGKLY EFASS	7.7e-27	243	78.3%
Q52Z19	<i>Pisum sativum</i>	Suppressor of Overexpression of Constans1	MVRGKTQMKRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVALIIFSPRGK	2.8e-30	276	98.1%
W9S5P5	<i>Morus notabilis</i>	MADS-box transcription factor 50	MVRGKTQMRRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVALIIFSPRGK	2e-31	273	96.2%
E1CKY5	<i>Vigna unguiculata</i>	Suppressor of overexpression of constans 1	MVRGKTQMRRIENATSRQVTFKRRN GLLKKAFELSVLCDAEVALIIFSPRGK	7.6e-30	273	96.2%
D1MDP7	<i>Vitis vinifera</i> (Grape)	Suppressor of overexpression of CO 1	MVRGKTQMRRIENATSRQVTFKRRN GLFKKAFELSVLCDAEVALIIFSPRGK	3.3e-29	269	94.3%

De maneira geral, o conteúdo CG das sequências gênicas do domínio MADS-box SRF em ambos os genes *SOC1* e *FUL* variou entre 42% e 51,4% (Tabela 5). Sabe-se que o par CG está ligado por três pontes de hidrogênio, enquanto o par AT é ligado por duas ligações de hidrogênio. A porcentagem do conteúdo GC pode revelar informações sobre a estabilidade desse gene. No entanto, não há dados na literatura sobre essa proposição. Neste sentido, não é possível afirmar que o conteúdo de CG no domínio do *MADS SRF* dos genes *FUL* e *SOC1* esteja ligado à termotolerância.

Tabela 5: Relação da quantidade de CG nas sequências gênicas *FUL* e *SOC1* para os genótipos de mamoeiro JS12, SS72//12, Cariflora, Calimosa e Golden.

Gene / Genótipo	% GC no domínio do MADS SRF
Gene Ful <i>Carica papaya</i> ( <i>Phytozome</i> )	42%
Gene Ful / Calimosa	51,14%
Gene Ful / Cariflora	42%
Gene Ful / SS72/12	43,83%
Gene Ful / Golden	43,34%
Gene Ful / JS12	42,46%
Gene SOC1 <i>Carica papaya</i> ( <i>Phytozome</i> )	42,98%
Gene SOC1/ Calimosa	46,05%
Gene SOC1/ Cariflora	47,36%
Gene SOC1/ SS72/12	45,61%
Gene SOC1/ Golden	45,17%
Gene SOC1/ JS12	46,49%

As sequências de aminoácidos para o domínio SRF das sequências do mamoeiro referentes ao gene *FUL* apresentaram algumas diferenças entre os genótipos avaliados. Para a análise das sequências de aminoácidos, além dos cinco genótipos em análise nesse trabalho e da sequência do banco de dados biológicos do *Phytozome*, utilizou-se uma sequência do banco *Uniprot*.

Para a classificação das possíveis diferenças, todas as sequências do gene *FUL* dos cinco genótipos foram comparadas com as sequências

depositadas do banco do *Phytozome* e *Uniprot*, denominadas *supercontig\_14.223* e *Carica*, respectivamente (Figura 6).

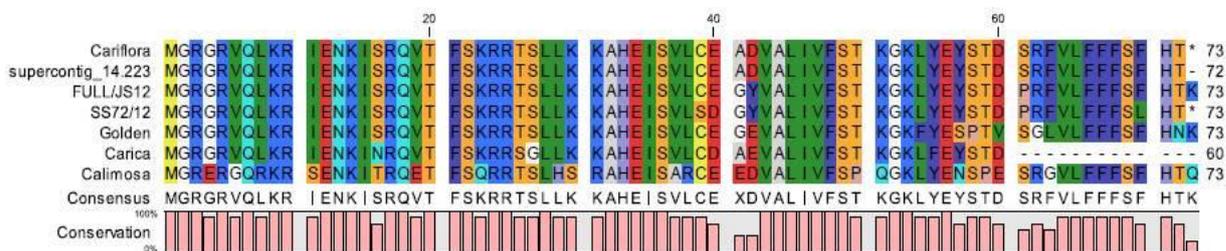


Figura 6: Alinhamento múltiplo entre todas as sequências de proteínas MADS SRF para o gene FUL. Podem ser visualizadas as sequências do domínio MADS dos genótipos Cariflora, JS12, SS72/12, *Golden*, Calimosa e as sequências *supercontig\_14.223* do banco de dados *Phytozome* e a sequência de aminoácidos *Carica* do banco de dados *Uniprot* evidenciado pelo programa *CLC Workbench8*.

A Figura 6 mostra que todas as sequências dos aminoácidos correspondem ao domínio SRF, com tamanho de aproximadamente 60 resíduos incluindo a Metionina inicial (ATG- códon iniciador). O genótipo Cariflora não apresentou diferença estrutural, sendo a sequência de aminoácidos idêntica à sequência depositada no *Phytozome*. Os genótipos JS12, SS72/12, *Golden* e Calimosa apresentaram substituições nas sequências dos aminoácidos. Algumas substituições foram conservativas sendo incorporado o novo aminoácido de natureza química similar ao anterior, diminuindo a chance de uma alteração grave na estrutura da proteína, porém, algumas dessas substituições foram não conservativas e, neste caso, o novo aminoácido tem natureza química bem diferente ao antigo, que pode levar a uma maior alteração na estrutura da proteína.

Há aminoácidos com cadeias laterais apolares como a alanina, a cisteína, a glicina, a valina, a leucina, a isoleucina, a prolina, a fenilalanina, o triptofano e a metionina. Aminoácidos polares neutros são um grupo de aminoácido que tem cadeias laterais polares eletricamente neutras (sem cargas) em pH neutro. Este grupo inclui a serina, a treonina, a tirosina, a glutamina e a asparagina. Aminoácidos polares ácidos são dois aminoácidos, o ácido glutâmico e o ácido aspártico, que têm grupos carboxila em suas cadeias laterais, além daquele presente em todos os aminoácidos. Há três aminoácidos polares básicos,

a histidina, a lisina e a arginina que têm cadeias laterais básicas, sendo, em todos eles a cadeia lateral é carregada positivamente em pH neutro ou perto dele.

O resíduo de aminoácido 42 do genótipo JS12 é a substituição de um ácido aspártico (D) por uma tirosina (Y), sendo uma substituição do tipo não conservativa, um aminoácido polar ácido por um aminoácido polar neutro.

O resíduo de aminoácido 39 do genótipo SS72/12 é a substituição de uma cisteína (C) por uma serina (S), sendo uma substituição do tipo não conservativa, um aminoácido de um resíduo apolar por um com polaridade neutra. O resíduo de aminoácido 42 do genótipo SS72/12 é a substituição de um ácido aspártico (D) por uma tirosina (Y), sendo uma substituição do tipo não conservativa, um aminoácido polar ácido por um aminoácido polar neutro. Essa substituição ocorre também no genótipo JS12.

O resíduo de aminoácido 54 e 57 do genótipo Golden é a substituição de uma leucina (L) por uma fenilalanina (F) e de uma tirosina (Y) por uma serina (S), respectivamente. Ambas as substituições são conservativas, pois as primeiras são proteínas de cadeia apolar e as segundas são proteínas com polaridade neutra, e o fato de a natureza química dos aminoácidos ser similar, diminui a chance de uma alteração grave na estrutura da proteína. O resíduo de aminoácido 58 é a substituição de uma serina (S) por uma prolina (P), sendo uma substituição do tipo não conservativa, um aminoácido neutro por um aminoácido apolar.

A sequência depositada do banco *Uniprot*, denominada *UniprotFUL* apresentou dois resíduos de aminoácidos diferentes da sequência de aminoácido depositada no *Phytozome*. O resíduo de aminoácido 16 é uma substituição conservativa do aminoácido polar neutro serina (S) por uma asparagina (N), também classificada como um aminoácido polar neutro. O resíduo de aminoácido 27 é uma substituição não conservativa do aminoácido polar neutro serina (S) por uma glicina (G), apolar. Estas substituições do gene *FRUITFULL* nos bancos de dados *Uniprot* e *Phytozome*, se refletem tanto na variação da porcentagem de identidade entre as sequências quanto possivelmente na função da proteína. Estas substituições podem ser assinaturas genéticas dos genes MADS-box em mamoeiro, devendo ser mais bem investigada.

O genótipo Calimosa apresentou número maior de substituições, ou seja, menor identidade, de 70% com a sequência depositada no banco *Phytozome*,

diferente das sequências de aminoácidos dos outros genótipos analisados, que foram de 93,4% a 100%. Esse fato por ser explicado por Calimosa ser um híbrido, que foi produzido na Caliman pela UENF, com o cruzamento de duas variedades pertencentes aos dois grupos heteróticos Solo e Formosa. O genótipo Calimosa tem um total de 17 substituições e dessas substituições os resíduos 6, 16, 19, 31, 37 e 57 foram classificados como substituições conservativas, e as substituições restantes foram não conservativas. O resíduo aminoácido 4 é a substituição de uma glicina (G) apolar por um aminoácido polar ácido, o ácido glutâmico (E). O resíduo de aminoácido 8 é a substituição de uma leucina (L) apolar por uma arginina (R) de caráter polar básico. O resíduo aminoácido 11 é a substituição de uma Isoleucina (I) com cadeia apolar por uma serina (S) de cadeia polar neutra. O resíduo de aminoácido 23 é a substituição de uma lisina (K) com polaridade básica por uma glutamina (Q) de caráter polar neutro. O resíduo de aminoácido 29 é a substituição de uma serina (S) de cadeia polar neutra por uma histidina (H) de polaridade básica. O resíduo de aminoácido 38 é a substituição de uma leucina (L) apolar por uma arginina (R) de polaridade básica. O resíduo de aminoácido 41 é a substituição de uma alanina (A) por um aminoácido de polaridade ácida o ácido glutâmico (E). O resíduo de aminoácido 50 é a substituição de uma treonina (T) com polaridade neutra pelo aminoácido apolar prolina (P). O resíduo de aminoácido 51 é a substituição de uma lisina (K) com polaridade básica para uma glutamina (Q) de polaridade neutra.

As sequências de aminoácidos para o domínio SRF das sequências do mamoeiro, referentes ao gene *SOC1*, apresentaram também algumas diferenças e, diferentemente do gene *FUL*, as sequências de aminoácidos quando comparadas com o banco *Uniprot* não apresentaram sequência similar para a espécie *Carica papaya* L.

A Figura 7 mostra que o domínio MADS- Box SRF para o gene *SOC1* apresenta tamanho de aproximadamente 60 resíduos, incluindo o códon iniciador ATG, referente ao aminoácido metionina, em todas as sequências dos aminoácidos. Para a classificação das possíveis diferenças, todas as sequências do gene *SOC1* foram comparadas com a sequência depositada do banco do *Phytozome*, denominada *supercontig\_18.166*.

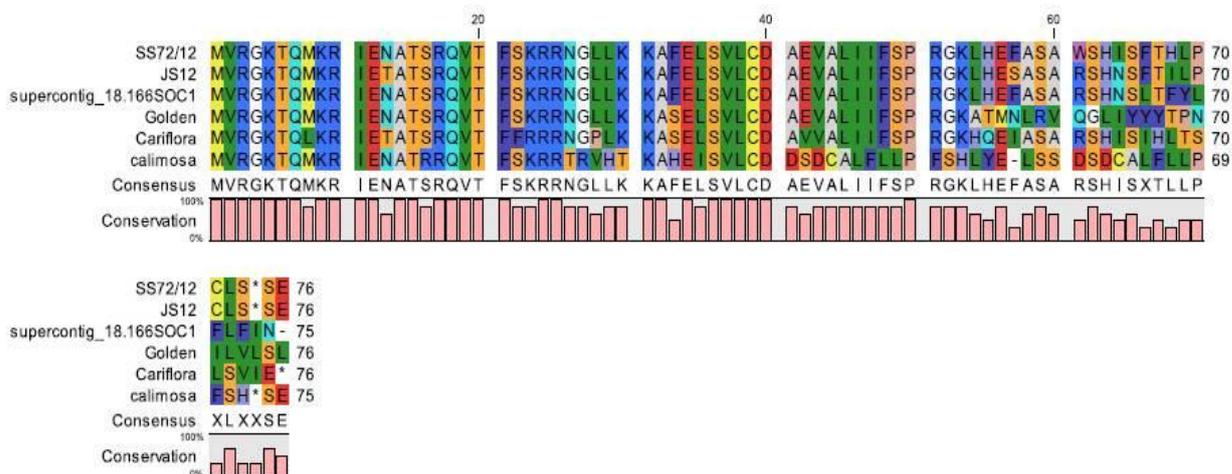


Figura 7- Alinhamento múltiplo, entre todas as sequências de proteínas MADS SRF, para o gene SOC1. Podem ser visualizadas as sequências do domínio MADS dos genótipos SS72/12, Cariflora, Golden, Calimosa e as sequências *supercontig\_18.166* do banco de dados *Phytozome*, evidenciado pelo programa CLC Workbench8.

O genótipo SS72/12 não apresentou diferença estrutural, sendo a sequência de aminoácidos idêntica à sequência depositada no *Phytozome*. Os genótipos JS12, Cariflora, Golden, e Calimosa apresentaram substituições nas sequências dos aminoácidos conservativas e não conservativas.

O resíduo de aminoácido 13 do genótipo JS/12 é a substituição de asparagina (N) uma por uma treonina (T), sendo uma substituição do tipo conservativa com aminoácidos de polaridade neutra. O resíduo de aminoácido 58 do genótipo JS/12 é a substituição de fenilalanina (F) pela serina (S), sendo uma substituição do tipo não conservativa de um aminoácido apolar para um com polaridade neutra.

O genótipo Golden apresentou uma substituição do tipo não conservativa no resíduo 33 da sequência do aminoácido apolar fenilalanina (F) para um com polaridade neutra serina (S) e substituições nos cinco últimos aminoácidos, sendo, dois do tipo conservativo nos resíduos 55 e 59. O resíduo de aminoácido 56 é a substituição de histidina (H) por uma treonina (T), de um aminoácido polar básico por um com polaridade neutra. O resíduo de aminoácido 57 é a substituição de um aminoácido polar ácido, o ácido glutâmico (E), por um apolar,

metionina (M). No resíduo de aminoácido 58 é a substituição do aminoácido apolar fenilalanina (F) por um neutro, asparagina (N).

O genótipo Cariflora apresentou nove substituições, sendo quatro conservativas nos resíduos 8, 13, 38 e 58. O resíduo de aminoácido 22 é a substituição de uma serina (S) por uma fenilalanina (F), um aminoácido polar neutro por um apolar. O resíduo de aminoácido 33 é a substituição de uma fenilalanina (F) por uma serina (S), um aminoácido apolar por um polar neutro. O resíduo de aminoácido 42 é a substituição de um ácido glutâmico (E) por uma valina (V), um aminoácido com polaridade ácida por um apolar. O resíduo de aminoácido 55 é a substituição de uma leucina (L) por uma histidina (V), um aminoácido apolar por um com polaridade básica. O resíduo de aminoácido 58 é a substituição de uma histidina (V) por uma glutamina (Q), um aminoácido apolar por um com polaridade neutra.

Como nas análises da sequência de aminoácidos do domínio SRF do gene *FUL*, as sequências de aminoácidos do gene *SOC1* do híbrido Calimosa apresentaram número maior de substituições quando comparadas com a sequência de aminoácido depositada no banco *Phytozome*, uma identidade baixa, de 61,6%, quando comparada com as outras sequências de mamoeiro analisadas, que ficaram entre 85%-100% de identidade, com a sequência do banco *Phytozome*. Do total de 23 substituições, dez foram classificadas como substituições conservativas, resíduos- 26, 28, 35, 44, 45, 46, 47, 53, 55 e 58 e as substituições restantes foram não conservativas.

O resíduo de aminoácido 16 do genótipo Calimosa é a substituição de serina (S) por uma arginina (R), sendo uma substituição do tipo não conservativa, de um aminoácido polar neutro por um polar básico. O resíduo de aminoácido 27 é a substituição de uma glicina (G) apolar por uma arginina (R) de polaridade neutra. O resíduo de aminoácido 29 é a substituição de uma leucina (L) apolar por uma histidina (R) com polaridade básica. O resíduo de aminoácido 30 é a substituição de uma lisina (K) com polaridade básica, por uma treonina (N) com polaridade neutra. O resíduo de aminoácido 33 é a substituição de uma fenilalanina (F) apolar, por uma histidina (H) com polaridade básica. O resíduo de aminoácido 41 é a substituição de uma alanina (A) apolar por uma com polaridade ácida, o ácido aspártico (D). O resíduo de aminoácido 42 é a substituição de aminoácido, com polaridade ácida, o ácido glutâmico (E), por uma serina (S), com

polaridade neutra. O resíduo de aminoácido 43 é a substituição de um aminoácido apolar, a valina (V), pelo ácido aspártico (D) com polaridade ácida. O resíduo de aminoácido 48 é a substituição de um aminoácido neutro, a serina (S), por uma leucina (L), apolar. O resíduo de aminoácido 51 é a substituição de uma arginina (R) polar neutra por uma fenilalanina (F) apolar. O resíduo de aminoácido 52 é a substituição de uma glicina (G) apolar por uma serina (N) polaridade neutra. O resíduo 57 fenilalanina (F) está ausente na sequência do aminoácido do genótipo Calimosa.

### 5.5 Filogenia molecular dos genes FUL e SOC1 em mamoeiro

As sequências dos genes MADS-box utilizadas para a inferência filogenética nesse trabalho tem somente um éxon. Por terem essa estrutura eles foram classificados como genes MADS-box do tipo I. Para a realização da inferência filogenética nesse trabalho foram utilizadas também as sequências do domínio MADS–SRF das espécies *Sinningia speciosa* da ordem *Lamiales*, *Morus notabilis* da ordem *Rosales*, *Vitis vinifera* da ordem *Vitales* e *Arabidopsis thaliana* da ordem *Brassicales*.

O resultado da busca dos genes FUL e SOC1 para a inferência filogenética com os genótipos de mamoeiro analisados neste trabalho mostrou algumas características interessantes. As sequências dos genes FUL e SOC1 da espécie *C.papaya*, presentes no *phytozome*, que têm o domínio MADS- box SRF quando contrastadas com o banco de dados curados *Uniprot*, geraram mais de 250 sequências com uma similaridade estatisticamente significativa, entre elas, uma sequência MADS tipo II para o gene *Fruitfull* (com um e-value de  $1.9e^{-30}$  e identidade de 88,3%) .

As sequências *phytozome\_FUL* e *Uniprot\_FUL* (*C.papaya*) foram alinhadas no *CLC Genomics Workbench 8* e observou-se que o domínio MADS-box altamente conservado SRF difere em cinco nucleotídeos ( Figura 8).

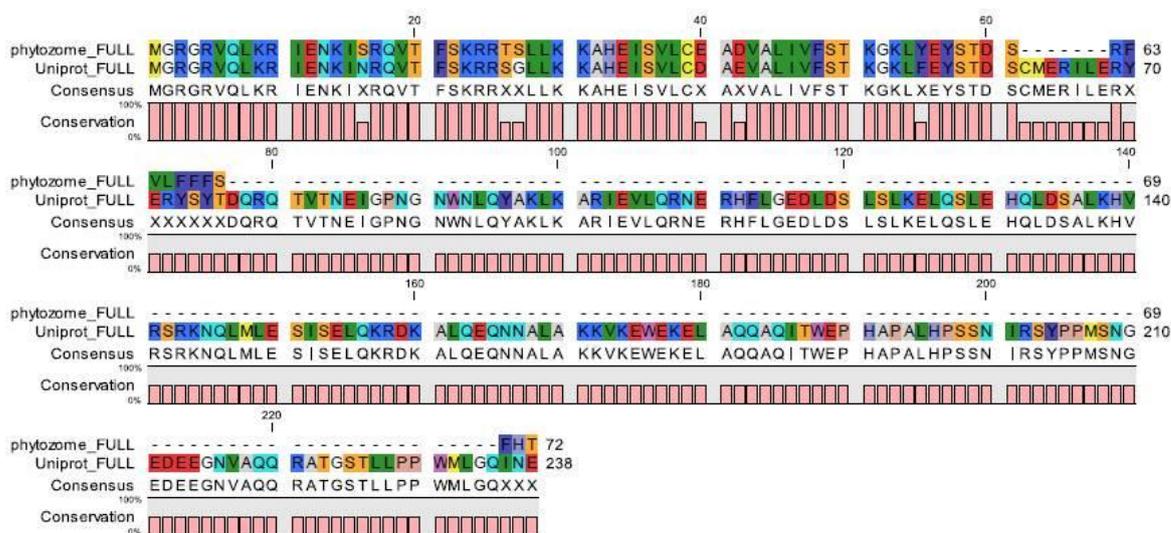


Figura 8: Visualização das seqüências do domínio MADS dos genótipos do banco de dados Phytozome, denominado *phytozome\_FUL* e uma seqüência depositada no banco uniprot, nomeada Uniprot\_FUL, evidenciado pelo programa CLC Workbench8.

A primeira diferença aparece no resíduo de aminoácido 16, na seqüência *phytozome\_FUL*, aparece uma serina (S), e na seqüência Uniprot\_FUL, aparece uma asparagina (N). No resíduo de aminoácido 26, na seqüência *phytozome\_FUL*, aparece uma treonina (T), e na seqüência Uniprot\_FUL, aparece uma serina (S). No resíduo de aminoácido 27, na seqüência *phytozome\_FUL*, aparece uma treonina (T), e na seqüência Uniprot\_FUL, aparece uma serina (S). No resíduo de aminoácido 40, na seqüência *phytozome\_FUL*, aparece um ácido glutâmico (E), e na seqüência Uniprot\_FUL, aparece um ácido aspártico (D). No resíduo de aminoácido 55, na seqüência *phytozome\_FUL*, aparece uma tirosina (Y), e na seqüência Uniprot\_FUL, aparece uma fenilalanina (F). As diferenças nos resíduos de aminoácidos 16, 26 e 40 foram conservativas, ou seja, os aminoácidos são de natureza química similar, diminuindo a chance de uma alteração grave na estrutura da proteína, porém, nos resíduos de aminoácidos 27 e 55, as diferenças foram não conservativas, neste caso, o novo aminoácido tem natureza química bem diferente do antigo, que pode levar a uma grave alteração na estrutura da proteína.

Ortólogos são genes cujas seqüências mais se assemelham á de um ancestral comum, mas que divergiram por especiação. Genes ortólogos mantêm a mesma função em duas espécies que divergiram. Os genes ortólogos FUL, das



considerados do tipo II pela presença de mais três domínios - intermediário (I), queratina (K) e região c-terminal (C) - além do domínio MADS-box SRF. O gene FUL da espécie *Morus Notabilis* (amora) é classificado como MADS tipo I, pois só apresenta o domínio SRF.

Os genes homólogos do clado D apresentaram similaridade, variando de 83.9% a 93.3%, com os genes FUL dos genótipos de mamoeiro do grupo interno SS72/12, JS12, Cariflora, Golden e Calimosa.

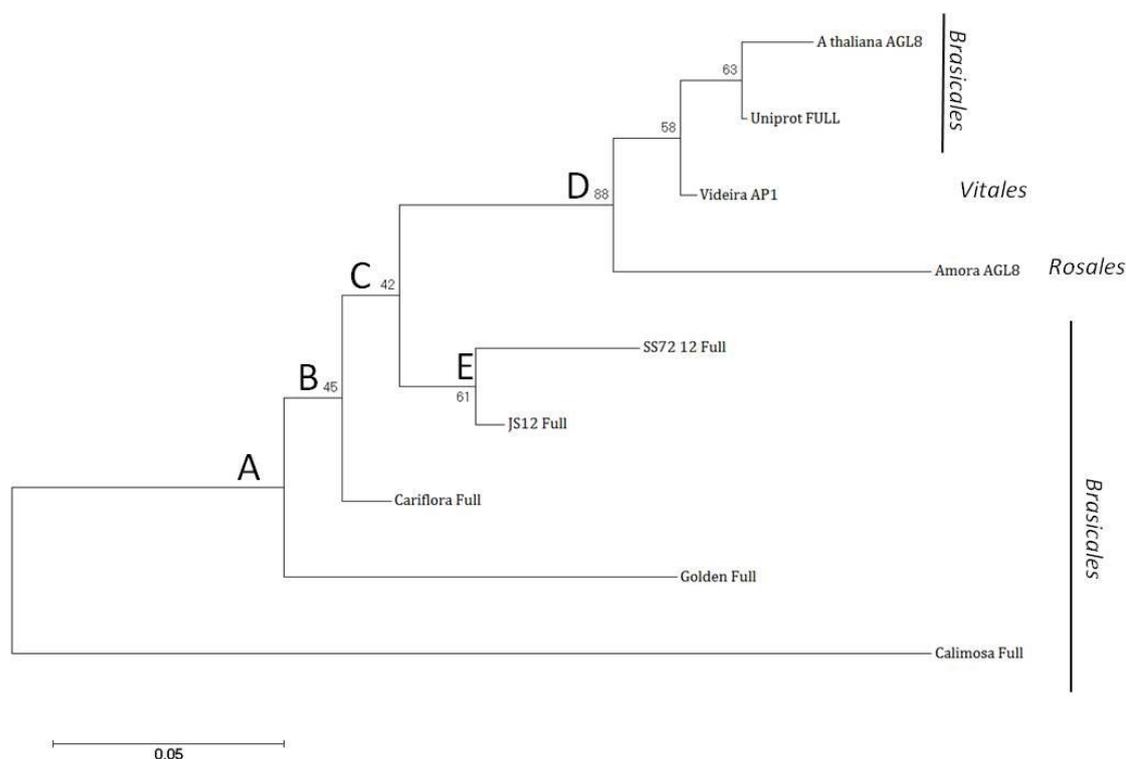


Figura 10: Árvore filogenética de sequências de proteínas do domínio SRF do gene MADS-box FUL, relativas aos genótipos de mamoeiro. Para uma análise comparativa foram usadas sequências proteicas do domínio SRF das espécies *Arabidopsis thaliana*, *Morus notabilis*, *Vitis vinifera* e uma sequência *Carica papaya* depositada no banco Uniprot. A árvore filogenética do gene FUL foi construída utilizando o programa MEGA6.

De acordo com Qingyi Yu (2008) o padrão de expressão do gene FUL em *Carica papaya* ocorre em flores e folhas dos três tipos sexuais, femininas masculinas e hermafroditas. Dentro órgãos florais específicos, a expressão foi detectada em sépalas, pétalas, e carpelos, mas não nos estames, um padrão que é típico para os genes EuFULI (Litt, 2007). Em *Arabidopsis*, o gene FUL/AGL8 é classificado como o gene que promove a identidade em estádios iniciais do

meristema floral em sinergia com APETALA 1 (AP1) e CAULIFLOWER (CAL). É requerido para a transição de um meristema inflorescência para um meristema floral e também para a padronização da divisão celular, expansão e diferenciação durante a morfogênese do fruto (Ferrándiz et al., 2000; Balanzà et al., 2014; Torti et al., 2012; Melzer et al., 2008).

Os limitados dados indicam que existem algumas funções que são comuns entre os parálogos, ou seja, identidade do meristema floral, mas outros que são exclusivos e nem mesmo talvez conservados dentro de uma família de genes, particularmente, identidade do perianto em euAP1 (Litt, 2007).

O gene no FUL dos genótipos JS12 e SS72/12 compartilham um ancestral comum com o clado D (Figura 10). O domínio altamente conservado SRF desses genótipos difere em dois resíduos, mas são diferenças conservativas, pois envolvem resíduos com a mesma característica química.

Calonje M. e colaboradores em 2004 identificaram dois genes da subfamília SQUA em videira, o gene *Vitis-FUL* e o *Vitis-AP1*. Análises de expressão puderam verificar que o gene VAP1 não apresentou padrão de expressão característico no desenvolvimento de sépalas e pétalas. *Vitis-FUL* e VAP1 apresentaram uma coexpressão no meristema reprodutivo, sendo o *Vitis-AP1* restrito à região do carpelo e o *Vitis-AP1*, a uma expressão no carpelo e estames.

A sequência Uniprot\_FUL divide o mesmo ancestral com a sequência Videira\_AP1 (homólogo *Vitis-AP1*) (Figura 9). A região MADS-Box com o domínio altamente conservado SRF difere somente em um resíduo de aminoácido. O resíduo 26 na sequência *Uniprot\_FUL* é uma serina (S) e na sequência Videira\_AP1 é uma treonina (T) (Figura 11). Serina e treonina são aminoácidos polares neutros, caracterizando uma diferença conservativa, diminuindo a chance de uma alteração grave na estrutura da proteína, podendo, assim, ter sua função biológica e processo evolutivo conservado em ambas as espécies.

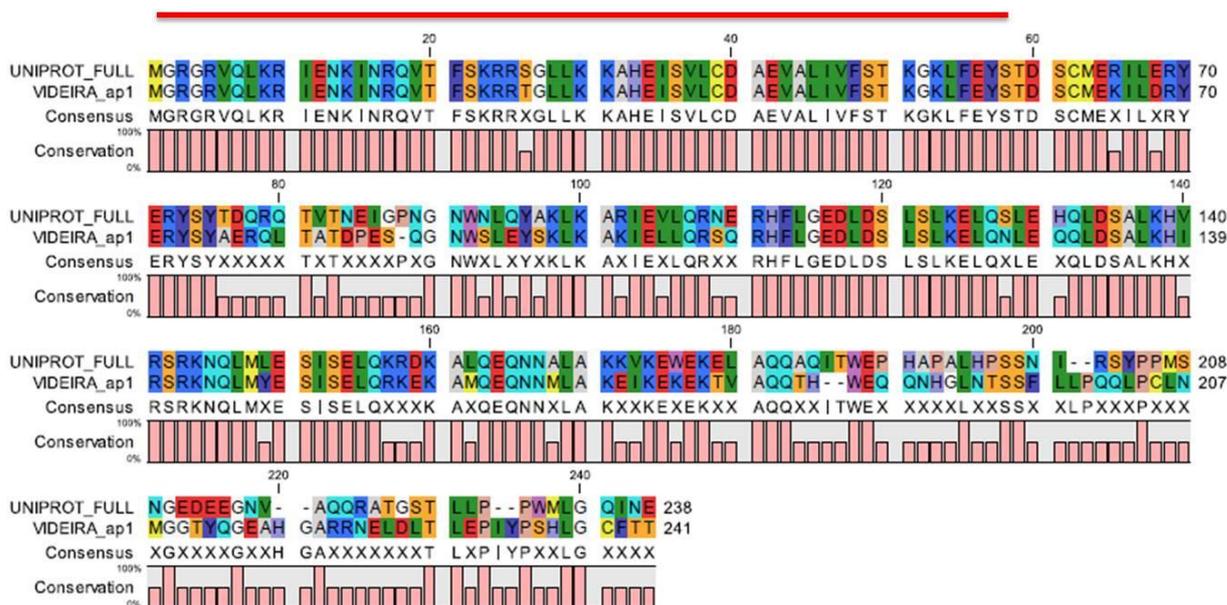


Figura 11: Alinhamento múltiplo entre todas as sequências de proteínas MADS SRF para o gene FUL. Podem ser visualizadas as sequências do domínio MADS dos genótipos do banco de dados Uniprot denominado Uniprot\_FUL e *Vitis*-AP1, evidenciado pelo programa CLC Workbench8.

Há trabalhos que relatam os fatores transcricionais MADS-box da classe B, *Fruitfull*, desempenhando um papel no tempo de floração em diversas espécies. O fator transcricional está ligado a várias vias do florescimento para promover a floração como - fotoperíodo, temperatura do ambiente e idade da planta. Estudos indicam o gene FUL atua de forma redundante na transição floral com outro fator transcricional - SOC1. O gene FUL e o gene SOC1 compartilham uma região regulatória, pois ambos são ativados pelo complexo FT-FD e reprimidos por outro fator transcricional, o gene Short Vegetative Phase (SVP).

O gene denominado SOC1/AGL20 em *Arabidopsis* é classificado funcionalmente como o gene que controla a floração. Esse gene pode integrar sinais especiais do ambiente, como fotoperíodo, vernalização e vias autônomas de indução floral, modular a expressão dos genes homeóticos da classe B (Apetala1/AP1, Fruitfull/FUL) e C (Agamous/AG, Sepalata/SEP) e quando associado com o gene AGL24, pode intervir no efeito da giberelina na floração em condições de dia curto, regulando a expressão de LFY. A habilidade que os sinais indutivos da floração têm de alterar o desenvolvimento da planta foi revelada ser dependente da competência do ápice meristemático em responder a esses sinais, envolvendo intimamente a alteração transcricional do fator LFY. O gene SOC1 e o

fator LFY apresentam funções independentes e se sobrepõem na determinação do tempo de floração e na iniciação floral (Araki, 2001; Simpson e Dean, 2002; Blázquez et al., 1997)

A árvore filogenética, Figura 12, compreendendo a homologia entre as sequências de proteínas do gene MADS-box SOC1 de *Arabidopsis thaliana*, *Vitis vinifera* (videira) e os genótipos de mamoeiro, foi dividida em grupos.

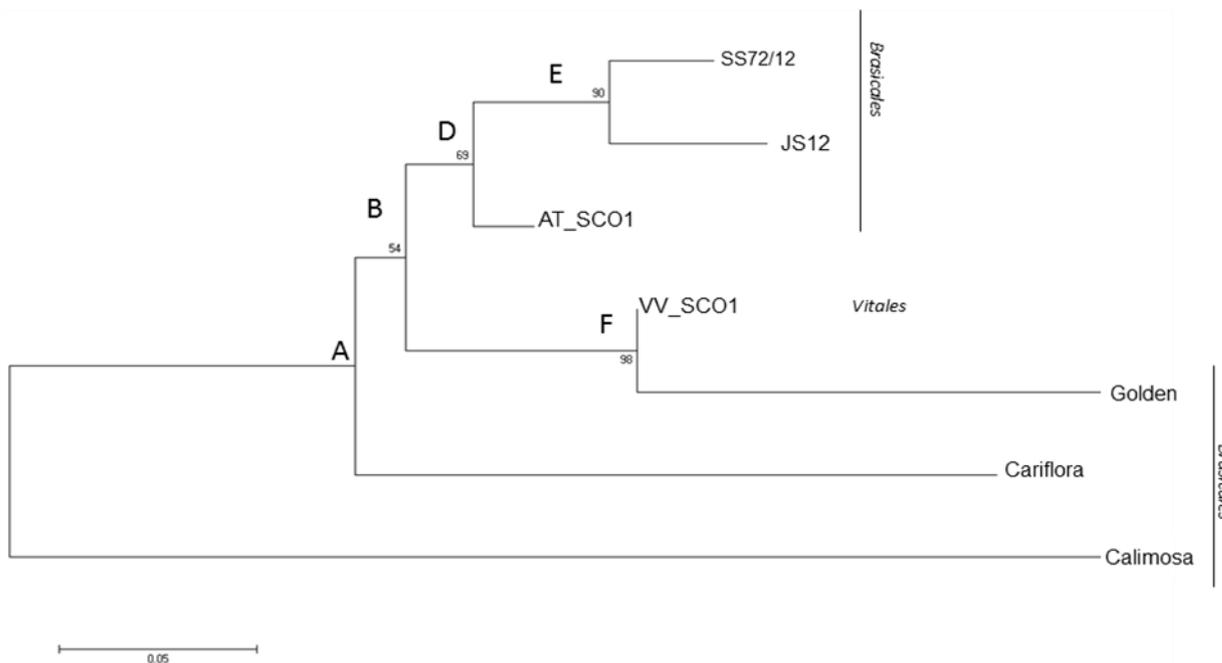


Figura 12: Árvore filogenética de sequências de proteínas do domínio SRF do gene MADS-box SOC1 relativos aos genótipos de mamoeiro. Para uma análise comparativa foram usadas sequências proteicas do domínio SRF das espécies *Arabidopsis thaliana* e *Vitis vinifera*. A árvore filogenética do gene FUL foi construída utilizando o programa MEGA6

As sequências gênicas dos genótipos de mamoeiro JS12 e SS72/12, homólogas ao gene SOC1 compartilham ancestralidade com o clado irmão D, formado pela sequência do domínio MADS-box SRF de *Arabidopsis thaliana* (AT\_SOC1). O domínio altamente conservado SRF do genótipo SS72/12 apresenta similaridade de 100% com a sequência do mamoeiro predita como sendo MADS-box SOC1.

SS72/12\_SOC1, quando comparado com a sequência do domínio SRF de *Arabidopsis* (AT\_SOC1), diferiu em um total de três resíduos, Figura 13, os resíduos de aminoácidos 44, 51 e 55. O resíduo de aminoácido 44 do genótipo

SS72/12\_SOC1 é uma alanina (A) de caráter apolar e no genótipo AT\_SOC1 é uma serina (S) com polaridade neutra. O resíduo de aminoácido 51 do genótipo SS72/12\_SOC1 é uma histidina (H) com polaridade básica e no genótipo AT\_SOC1 é uma tirosina (Y) com polaridade neutra. As duas diferenças descritas foram não conservativas, o que pode levar a uma alteração na estrutura da proteína e possivelmente na sua função.

O resíduo de aminoácido 55 do genótipo SS72/12\_SOC1 é uma arginina (R) e no genótipo AT\_SOC1 é uma lisina (K), ambas de polaridade básica, uma diferença conservativa, e isso diminui a chance de uma alteração grave na estrutura da proteína.

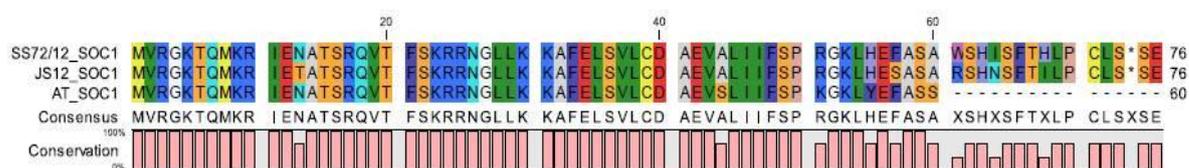


Figura 13: Alinhamento múltiplo entre todas as sequências de proteínas MADS SRF para o gene SOC1. Podem ser visualizadas as sequências do domínio MADS dos genótipos de mamoeiro SS72/12 e JS12, evidenciado pelo programa CLC Workbench8.

O clado F é formado pelos ortólogos VV\_SOC1 e Golden (Figura12). De acordo com Sreekantan L. e Thomas M. R. (2006), o padrão de expressão do gene SOC1 em *Vitis vinifera* foi baixo em raiz, folhas jovens, gavinhas em fase de desenvolvimento juvenil e não foi detectada em flores e frutos. A expressão do gene SOC1 em *Vitis vinifera* foi detectada no período muito inicial do meristema de inflorescência e nas brácteas que as envolviam e, à medida que os órgãos florais iam se desenvolvendo, a expressão do gene diminuía.

Foi detectado um alto nível de expressão relativa do gene SOC1 nas gemas axilares no momento em que a videira iniciava o processo de formação dos brotos da inflorescência, assim, pode-se sugerir que o gene SOC1 está fortemente ligado à indução floral em *Vitis vinifera*.

Golden\_SOC1 quando comparado com a sequência do domínio SRF de *Vitis vinifera* (VV\_SOC1), diferiu em um total de três resíduos (Figura 14). As diferenças nos resíduos de aminoácidos 9, 29 foram conservativas, sendo os aminoácidos de natureza química similar, diminuindo a chance de uma alteração grave na estrutura da proteína. O resíduo de aminoácido 55 apresenta diferença natureza química similar, isso diminui a chance de uma alteração grave na

estrutura da proteína. O resíduo de aminoácido 55 apresenta diferença no caráter químico, o que pode levar a uma alteração na estrutura da proteína e, possivelmente, na sua função.

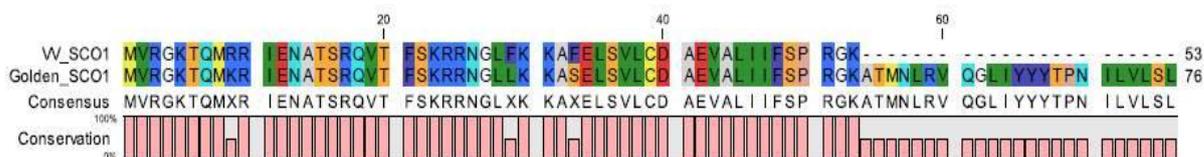


Figura 14: Alinhamento múltiplo entre todas as sequências de proteínas MADS SRF para o gene SOC1. Podem ser visualizadas as sequências do domínio MADS dos genótipos de mamoeiro *Golden\_SOC1* e uma sequência do banco de dados Uniprot da espécie *Vitis vinifera*, evidenciado pelo programa CLC Workbench8.

Tanto o gene SOC1 e quanto o FUL dos genótipos de mamoeiro analisados têm domínio MADS do tipo I. Vários fatores transcricionais MADS-box tipo I têm sido relatados por desempenhar papéis importantes no desenvolvimento reprodutivo em *Arabidopsis*. Todos os genes MADS do tipo I, submetidos à análise de expressão, mostraram ser ativos no gametófito feminino e no desenvolvimento de sementes, indicando que a maior parte da subfamília de genes MADS-box do tipo I está envolvida no desenvolvimento reprodutivo em *Arabidopsis*. Curiosamente, a expressão foi predominantemente observada na célula central, antípodas e endosperma (Bemer M. et al., 2010).

A floração em plantas perenes, tais como *Vitis vinifera* e *Carica papaya*, difere significativamente no desenvolvimento quando comparadas com espécies anuais como *Arabidopsis thaliana* (Boss et al., 2003). O padrão de desenvolvimento floral se reflete diretamente na expressão relativa dos genes reguladores, principalmente aqueles responsáveis pelo desenvolvimento floral, como os genes MADS-box.

## 6. CONCLUSÕES E PERSPECTIVAS FUTURAS

Principais conclusões tendo como base a análise dos dados são:

- Os primers construídos nesse trabalho se amplificaram em uma região específica do genoma e exibiram especificidade satisfatória. Os primers específicos para o gene SOC1 amplificaram um produto de 580pb, e os primers específicos para o gene FUL amplificaram um produto de 840pb. O tamanho do fragmento gerado após a reação de PCR está condizente com o tamanho do fragmento descrito na construção pelo programa *primer 3 plus*.
- Foi possível, com o auxílio do CodonCode, formar contigs com as sequências para cada genótipo analisado. Ambos contigs *Forward* e *Reverse* de cada gene montados formou um único *contig* de sequências nucleotídicas com o tamanho de 580pb para o possível gene SOC1 e contig com 840 pb para o possível gene FUL em cada genótipo.
- Foi possível identificar com o auxílio do banco de dados biológicos on-line *Pfam*, os domínios MADS–box de todos os dez *contigs*.

- A sequência de aminoácidos dos CDS referentes aos dez contigs montados no sequenciamento, comparados com o banco de dados *Uniprot* e *Swiss-Prot*, apresentou conservação in silico do domínio SRF. Para o gene FUL, o genótipo Cariflora apresentou 100% de conservação in silico com a sequência de aminoácidos depositada no *Phytozome*, e os genótipos JS12, SS72/12, Golden e Calimosa apresentaram algumas substituições não conservativas e conservativas nas sequências dos aminoácidos. Para o gene SOC1, o genótipo SS72/12 apresentou 100% de conservação in silico com a sequência de aminoácidos depositada no *Phytozome*. Os genótipos JS12, Cariflora, Golden e Calimosa apresentaram substituições nas sequências dos aminoácidos conservativas e não conservativas.
- Os experimentos de bioinformática irão auxiliar o direcionamento dos experimentos de bancada, podendo validar por qPCR os resultados obtidos pela bioinformática. Para a inferência funcional dos genes Tipo I de mamoeiro presente neste trabalho, as análises de expressão gênica precisam ser integradas.
- Verificou-se nas análises de busca por homologia no bando de dados do *Carica papaya* que o gene *FUL* apresenta uma possível duplicação adicional do gene com o domínio MICK. A sequência de aminoácidos do gene FUL Tipo II, depositada no phytozome, não apresenta conserção 100% in silico com as sequências depositadas no banco *Uniprot* sendo que ambas as sequências apresentam uma alta similaridade com o gene Ful tipo I, depositado no phytozome.

Para a padronização das análises, é essencial fazer uma reanotação do genoma do *Carica papaya L.* e também uma mineração de dados de sequências preditas referentes à família MADS-box para uma inferência funcional mais eficiente. Com uma mineração de dados adequada pode-se optar por uma análise de predição computacional de interações de proteína- proteína em proteomas preditos de *Carica Papaya L.*, e isso pode representar uma contribuição importante para a predição de função dos genes MADS-box no desenvolvimento

floral e quais são os genes que sofrem influência sazonal alterando o padrão de expressão e conseqüentemente a morfologia floral.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Adamczyk, B.J. & Fernandez, D.E., (2009) MIKC\* MADS-domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiology*, 149:1713-1723.
- Adebivi A., Adaikan P.G. and Prasad R.N., (2002) Papaya (*Carica papaya*) consumption is unsafe in pregnancy, fact or fable? Scientific evaluation of a common belief in some parts of Asia using a rat model. *Journal Nutrition*, 88(2), 199-203.
- Ahuja I., De Vos R.C., Bones A.M., Hall R.D. (2010) Plant molecular stress responses face climate change. *Trends Plant Sci.* 15, 664.
- Alvarez-Buylla E.R., Pelaz S., Liljegren S.J., Gold S.E., Burgeff C., Ditta G.S., Ribas de Pouplana L., Martínez-Castilla L., Yanofsky M.F.(2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. USA* 97: 5328–5333.
- Alvarez-Buylla E.R., Liljegren S.J., Pelaz S., Gold S.E., Burgeff C., Ditta G.S., Vergara-Silva F. and Yanofsky M.F., (2000) MADS-box gene evolution beyond flowers: Expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J* 24:457-466.

- Araki T., (2001) Transition from vegetative to reproductive phase. *Curr. Opin. Plant Biol.*, 4, 63–68.
- Argout X, et al., (2011) The genome of *Theobroma cacao*. *Nat Genet.*;43(2):101-108.
- Awada M., and Ikeda W.S., (1957) Effects of water and nitrogen application on composition, growth, sugar in fruit, yield and sex expression of the papaya plants (*Carica papaya* L.). Hawaii Agricultural Experiment Station Technical Bulletin 33, University of Hawaii, Honolulu.
- Awada M., (1958) Relationships of minimum temperature and growth rate with sex expression of *papaya* plants (*Carica papaya* L.). Hawaii Agricultural Experiment Station Technical Bulletin, v. 38, University of Hawaii, Honolulu.
- Badillo V.M., (2000) *Carica* vs *Vasconcella* St.-Hil. (Caricaceae) con la rehabilitación de este último. *Ernstia* 10:74–79.
- Badillo, V.M. (1971) Monografía de la familia Caricaceae. Asociación de profesores, Universidad Central de Venezuela, Maracay.
- Balanzà V., Martínez-Fernández I., Ferrándiz C., (2014) Sequential action of FRUITFULL as a modulator of the activity of the floral regulators SVP and SOC1. *J. Exp. Bot.* 65:1193
- Bemer M., Heijmans K., Airoidi C., Davies B., Angenent G.C., (2010) An atlas of type I MADS box gene expression during female gametophyte and seed development in *Arabidopsis*. *Plant Physiol.* 154(1):287-300.
- Becker A., Theissen G., (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464-489.

- Blázquez M.A., Soowal L.N., Lee I., Weigel D., (1997) LEAFY expression and flower initiation in *Arabidopsis*. *Development*. 124(19):3835-44.
- Blázquez M.A., Ferrándiz C., Madueño F., Parcy F., (2006) How floral meristems are built. *Plant Mol Biol*; 60(6):855-70
- Beilstein M.A., Nagalingum N.S., Clements M.D., Mancheste S.R., Mathews S., (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States*. 127 (43): 18724-18728.
- Bodhankar S.L., Garg S.K and Mathur V.S. (1974) Antifertility screening of plants. Effect of five indigenous plants on early pregnancy in female albino rats. *Indian J Med Res*, 62(6): 831-837.
- Borner R., Kampmann G., Chandler J., Gleißner R., Wisman E., Apel K. and Melzer S. (2000) A MADS domain gene involved in the transition to flowering in *Arabidopsis*. *Plant J*. 24, 591–599.
- Boss P.K., Buckeridge E.J., Poole A., Thomas M.R. (2003) New insights into grapevine flowering. *Funct Plant Biol* 30 593–606.
- Boss P.K., Bastow R.M., Mylne J.S., Dean D. (2004) Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell* 16: S18–S31.
- Bose B.C., Saifi A.Q., Vijayvargiya R. and Bhagwat A.W., (1961) Pharmacological study of *Carica papaya* seeds with special reference to its anthelmintic action. Preliminary report, *Indian J Med Sci*, 15(1), 888-892.
- Coppens d'Eeckenbrugge G., Restrepo M.T., Jiménez D., (2007) Morphological and isozyme characterization of common papaya in Costa Rica. *Acta hort.*740, 109–120.

- Carvalho F. and Renner S.S., (2012) A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. *Mol. Phylogenet. Evol.* 65(1): 46-53.
- Calonje M., Cubas P., Martínez-Zapater J.M., Carmona M.J., (2004) Floral meristem identity genes are expressed during tendril development in grapevine. *Plant Physiol* 135: 1491–1501
- Calzada F., Yepez-Mulia L. and Tapia-Contreras A., (2007) Effect of Mexican medicinal plant used to treat trichomoniasis on *Trichomonas vaginalis* trophozoites. *J Ethnopharmacol* 113(2): 248-251.
- Causier B., Castillo R., Zhou J., Ingram R., Xue Y., Schwarz-Sommer Z., Davies B., (2005) Evolution in action: following function in duplicated floral homeotic genes. *Curr Biol.*15:1508–1512.
- Charlesworth B., Charlesworth D., (1978) A model for the evolution of dioecy and gynodioecy. *Am Nat*, 112: 975–997.
- Chen Y.C., Eisner J.D., Kattar M.M., Rassouljian-Barrett S.L., Lafe K., Bui U., Limaye A.P., Cookson B.T., (2001) Polymorphic internal transcribed spacer region 1 DNA sequences identify medically important yeasts. *J Clin Microbiol.* 39(11): 4042-51.
- Colunga-García M.P. and Zizumbo V.D., (2004). Domestication of plants in Maya lowlands. *Economic Botany* 58 (Supplement), pp: S101–S110.
- Conover R.A., Litz R.E., (1979) Breeding papayas tolerant to papaya ringspot virus. In: *Tropical Region, American Society for Horticultural Science, Homestead, Florida. Proceedings.* p.155-157.
- Ctenas M.L.B. and Ctenas A.C., (2000) Quast, D. *Frutas das terras brasileiras.* São Paulo: C2 Editora, 157.

- Damasceno Junior P.C., et al. (2010) Meiotic behavior of *Carica papaya* and *Vasconcellea monoica*. *Caryologia*, v. 63, p. 229-236.
- Dantas J.L.L., Castro Neto M.T., (2000) Aspectos botânicos e fisiológicos. In: Mamão produção: aspectos técnicos, Brasília:Embrapa Comunicação para Transferência de Tecnologia, p.11-14.
- Dantas J.L.L, Dantas A.C.V.L., Lima J.F., (2002) Mamoeiro, in: Bruckner, C. H. Melhoramento de fruteiras tropicais. UFV. Viçosa, MG. p.309-349.
- Denoeud F., et al. (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345: 1181-1184.
- De Oviedo G.F., (1959) Historia General y Natural de las Indias. JP de Tudela. Ediciones Atlas, Madrid.
- De Bodt S., Raes J., Van de Peer Y., Theissen G. (2003) And then there were many: MADS goes genomic. *Trends Plant Sci* 8 475–483.
- Dornelles A.L.C., (2008) Origem e evolução de plantas cultivadas. Embrapa Informação Tecnológica, 909 p.499-505.
- Dunning A.M., Healey C.S., Pharoah P.D., Foster N.A., Lipscombe J.M., Redman K.L., Easton D.F., Day N.E. and Ponder B.A., (1998) No association between a polymorphism in the steroid metabolism gene CYP17 and risk of breast cancer. *Br J Cancer* 77:2045-2047.
- Edger P.P., Pires J.C., (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* 17, 699–717.
- Edger P.P., Heidel-Fischer H.M., Bekaert M., et al. (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the*

*National Academy of Sciences of the United States of America*. 112 (27):8362-8366.

El Moussaoui A., Nijs M., Paul C., Wintjens R., Vincentelli J., Azarkan M., Looze Y. (2001). Revisiting the enzymes stored in the laticifers of *Carica papaya* in the context of their possible participation in the plant defence mechanism. *Cell Mol Life Sci* 58(4):556–570.

Ferrándiz C., Gu Q., Martienssen R., Yanofsky M.F., (2000) Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development* 127: 725-734.

FAOSTAT (2015) Food and Agriculture Organization of the United Nations Database. <http://www.apps.fao.org> . Accessed December 2015.

Felsenstein J., (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, v. 39, p. 783-791.

Freeling M., (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol.* 60:433–453.

Fuentes G. and Santamaría J. M., (2014) Papaya (*Carica papaya* L.): Origin, Domestication, and Production 10.1007/978-1-4614-8087-7\_1 pp 3-15

Gramzow L., Ritz M.S., Theissen G., (2010) On the origin of MADS-domain transcription factors. *Trends Genet.* 26: 149–153

Glenn T.C., (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 11, 759-769.

Gschwend A.R., Wai C.M., Zee F., Arumuganathan A.K. and Ming R. (2013) Genome size variation among sex types in dioecious and trioecious *Caricaceae* species. *Euphytica* 189: 461–469.

- Gramzow L., Theißen G., (2013) Phylogenomics of MADS-Box Genes in Plants - Two Opposing Life Styles in One Gene Family. *Biology (Basel)* 12;2(3):1150-64.
- Hart S.M., (2002) Modulation of nuclear receptor dependent transcription. *Biological Research*, 35:295-303.
- Hayes T.E., Sengupta P., Cochran B.H., (1998) The human c-fos serum response factor and the yeast factor GRM/PRTF have related DNA-binding specificities. *Gene Dev* 2:1713-1722.
- Harlan J.R., (1971) Agricultural origins: centers and non-centers. *Science* 174:468–474
- Hennig W., (1966) *Phylogenetic Systematics*. University of Illinois Press, Urbana
- Henschel K., Kofuji R., Hasebe M., Saedler H., Münster T., Theißen G., (2002) Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. *Mol Biol Evol.* 19:801–814.
- Hofmeyr J.D.J., (1939) Genetical studies of *Carica papaya* L., Africa Dept. Agric. For. Sci. Bul., 187: 1-64.
- IBGE (2015) Produção Agrícola Municipal 2013. Disponível em: <<http://www.ibge.gov.br/estadosat/>> Acesso em: 25 de Dezembro de 2015.
- IPGRI (1993) *Diversity of development*. Rome: International Board for Plant Genetic Resources.
- International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome *Nature* 436:793-800
- Jacob F. & Monod J., (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318–356.

- Jones D.T., Taylor W.R., Thornton J.M., (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 8(3):275-82.
- Kapoor M., Garg S.K. and Mathur V.S., (1974) Antioviulatory activity of five indigenous plants in rabbits, *Indian J Med Res*, 62(8),1225-1227.
- Krishnakumari M.K. and Majumder S.K., (1960) Studies on anthelmintic activities of seeds of *Carica papaya* Linn. *Ann Biochem Exp Med*, 20(1): 551-556.
- Kaufmann K., Melzer R., Theissen G., (2005) MIKC-type MADS-domain proteins: Structural modularity, protein interactions and network evolution in land plants. *Gene* 347: 183–198.
- Khan M.R., Hu J., Ali G.M., (2012) Reciprocal loss of CArG-Boxes and Auxin response elements drives expression divergence of MPF2-like MADS-box genes controlling calyx inflation PLoS ONE 7: e42781.
- Kim S., Yoo M., Albert V.A., Farris J.S., Soltis P.S., and Soltis D.E., (2004) Phylogeny and diversification of B-function genes in angiosperms: Evolutionary and functional implications of a 260-million year old duplication. *Am. J. Bot.* 91: 2102–2118.
- Krishna K.L., Paridhavi M. and Patel J.A., (2008) Review on nutritional, medicinal and pharmacological properties of Papaya (*Carica papaya* Linn.) *Natural Product Radiance*, Vol. 7(4), pp.364-373.
- Kumar L., Abraham A. and Srinivasan V., (1945) The cytology of *Carica papaya* Linn. *Indian J. Agric. Sci.* 15, 242–253.
- Lawton-Rauh A., (2003) Evolutionary dynamics of duplicated genes in plants. *Mol Phylogenet Evol.* 29(3):396-409.
- Laux T., Mayer K.F.X., Berger J., and Jürgens G., (1996) The WUSCHEL gene is required for shoot and floral meristem integrity *Development* 119, 397–418. in *Arabidopsis*. *Development* 122, 87–96.

- Litt A., (2007) An evaluation of A-function: evidence from the *APETALA 1* and *APETALA 2* gene lineages. *Int. J. Plant Sci* 168: 73–91
- Liu Z., Moore P.H., Ma H., Ackerman C.M., Ragiba M., Yu Q., Peari H.M., Kim M. S., Chariton J.W., Stiles J.I., Zee F.T., Andrew A.H., Ming R., (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* 427: 348-352.
- Li D., Liu C., Shen L., Wu Y., Chen H., Robertson M., Helliwell C.A., Ito T., Meyerowitz E., Yu H., (2008) A repressor complex governs the integration of flowering signals in *Arabidopsis*. *Developmental Cell* 15, 110–120.
- Melzer S., Lens F., Gennen J., Vanneste S., Rohde A., Beeckman T., (2008) Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nat Genet* 40: 1489–1492
- Ming R., (2008) Low X/Y divergence in four pairs of papaya sex-linked genes. *Plant Journal*. 53(1):124-32.
- Magallon S., Crane P.R. and Herendeen P.S., (1999) Phylogenetic pattern, diversity, and diversification of eudicots. *Ann. Mo. Bot. Gard.* 86 297–372.
- Messenguy F., Dubois E., (2003) Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* 316, 1-21.
- MDIC – Ministério do Desenvolvimento, Indústria e Comércio. Secretaria de Comércio Exterior (SECEX). Exportações Brasileiras 2013-2015. Disponível em: <[www.desenvolvimento.gov.br/sitio/secex](http://www.desenvolvimento.gov.br/sitio/secex)>. Acesso em: 16/12/2015.
- Mondragon-Palomino M. and Trontin C., (2011) High time for a roll call: gene duplication and phylogenetic relationships of TCP-like genes in monocots. *Ann. Bot.* 107, 1533–1544.

- Müller B. & Grossniklaus U., (2010) Model organisms-A historical perspective. *J. Proteomics* 73: 2054–2063.
- Ming R., Yu Q., Moore P.H., ( 2007) Sex determination in papaya. *Semin Cell Dev Biol.* 18(3):401-8.
- Ming R., Hou S., Feng Y., Yu Q., Dionne-Laporte A., Saw J.H., Senin P., Wang W., Ly B.V., Lewis K.L., et al., (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus) *Nature* 452:991–996.
- Ming R., Bendahmane A., Renner S.S., (2011) Sex chromosomes in land plants. *Annu Rev Plant Biol.* 62:485-514.
- Nam J., Kim J., Lee S., An G., Ma H., Nei M., (2004) Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc Natl Acad Sci USA* 101 1910–1915
- Natsume S., Takagi H., Shiraishi A., Murata J., Toyonaga H., Patzak J., Takagi M., Yaegashi H., Uemura A., Mitsuoka C., Yoshida K., Krofta K., Satake H., Terauchi R., Ono E., (2015) The Draft Genome of Hop (*Humulus lupulus*), an Essence for Brewing. *Plant Cell Physiol.* 56(3):428-41.
- Nguyen T.T., Shaw P.N., Parat M.O., Hewavitharana, A.K., (2013) Anticancer activity of *Carica papaya*: a review. *Molecular Nutrition Food Res.*; 57 (1): 153-64
- Osato J.A., Santiago L.A., Remo G.M., Cuadra M.S. and Mori A., (1993) Antimicrobial and antioxidant activities of unripe papaya, *Life Sci*, 53(17), 1383-1389.
- Olson M.E., (2002) Combining data from DNA sequences and morphology for a phylogeny of Moringaceae (Brassicales). *Syst. Bot.* 27, 55–73.

- Pajoro A., Madrigal P., Muino J.M., et al., (2014) Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biology* 15, R41.
- Panse T.B. and Paranjpe A.S., (1943) Isolation of carpasemine from papaya seeds. *Proc Indian Acad Sci*, 18 A: 140.
- Parenicova L., de Folter S., Kieffer M., Horner D.S., Favalli C., Busscher J., Cook H.E., Ingram R.M., Kater M.M., Davies B., et al., (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 15, 1538-1551.
- Pope K.O., Pohl M.E., Jones J.G., Lentz D.L., vonNagy C., Vega F.J., (2001) Origin and environmental setting of ancient agriculture in the lowlands of Mesoamerica. *Science* 292:1370–1373.
- Portal Toda Fruta. Disponível em: <http://www.todafruta.com.br>. Acesso em 18 de janeiro de 2016.
- Praveen G.B. and Namita S., (2001) In vitro Antimalarial activity of extracts of three plants used in the traditional medicine of India. *Am J Trop Med Hyg* 65(4): 304-308.
- Robles P. and Pelaz S. (2005) Flower and fruit development in *Arabidopsis thaliana*. *Int J Dev Biol* 49: 633-643.
- Stein, L. (2001) Genome annotation: From sequence to biology. *Nat. Rev. Genet.* 2: 493-503.
- Rozen S. and Skaletsky H., (2000) Primer3 on the WWW for general users and for biologist programmers, *Methods Mol Biol*, 132, 365-386.

- Richards T.A., Soanes D.M., Foster P.G., Leonard G., Thomton C.R., Talbot N.J., (2009) Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell*. 21:1897–1911.
- Schwarz-Sommer Z., Huijser P., Nacken W., Saedler H., Sommer H., (1990) Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science*, 250, 931-936.
- Storey W.B., (1941) The botany and sex relations of papaya. Hawaii Agricultural Experiment Station Bulletin 87, 5-22.
- Shore P., Sharrocks A.D., (1995) The MADS- box family on transcriptional factors. *Eur J Biochem* 229: 1-13.
- Schmutz J., Cannon S. B., Schlueter J., Ma J., Mitros T., Nelson W., et al., (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183.
- Sreekantan L. and Thomas M.R., (2006) *VvFT* and *VvMADS8*, the grapevine homologues of the floral integrators *FT* and *SOC1*, have unique expression patterns in grapevine and hasten flowering in *Arabidopsis*. *Funct Plant Biol* 33 1129–1139
- Samach A., Onouchi H., Gold S.E., Ditta G.S., Schwarz-Sommer Z., Yanofsky M.F., Coupland G., (2000) Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science* 288,1613–1616.
- Stevens P.F., (2001) Angiosperm Phylogeny Website. Version 12, July. <http://www.mobot.org/MOBOT/research/APweb/>
- Simpson G.G. and Dean C., (2002) *Arabidopsis*, the Rosetta stone of flowering time? *Science* 296, 285–289.
- Sung S., Amasino R.MR., (2004) Vernalization and epigenetics: how plants remember winter. Volume 7, Issue 1, Pages 4–10.

- Shannon S., and Meeks-Wagner D.R., (1991) A mutation in the *Arabidopsis* TFL7 gene affects inflorescence meristem development. *Plant Cell* 3, 877-892.
- Sarala N. and Paknikar S.S., (2014) Papaya Extract to Treat Dengue: A Novel Therapeutic Option? *Ann Med Health Sci Res.* May-Jun; 4(3): 320–324.
- Shaogui Guo et al., (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics* 45, 51–58.
- Singh R et al., (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* 500:335-9.
- Soltis D.E., Gitzendanner M.A., Soltis P.S., (2007) A 567-taxon data set for angiosperms: The challenges posed by Bayesian analyses of large data sets. *International Journal of Plant Sciences* 168: 137–157.
- Thornton K. and Long M., (2002) Rapid divergence of gene duplicates on the *Drosophila melanogaster* X chromosome. *Mol Biol Evol.*19(6):918-25.
- Torti S., Fornara F., Vincent C., Andrés F., Nordström K., Göbel U., Knoll D., Schoof H., Coupland G. (2012) Analysis of the *Arabidopsis* shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. *Plant Cell* 24: 444–462
- Theissen G., Kim J.T., Saedler H., (1996) Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene families in the morphological evolution of eukaryotes. *J Mol Evol.* 43:484–516
- Theissen G., Becker A., di Rosa A., Kanno A., Kim J.T., Munster T., Winter K.U., Saedler H., (2000) A short history of MADS-box genes in plants. *Plant Mol. Biol.* 42, 115-149.
- Tamura K., Dudley J., Nei M. & Kumar S., (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599.

- Tuskan G.A., et al., (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793), 1596-1604.
- Tan F. and Swain S.F., (2006) Genetics of flower initiation and development in annual and perennial plants. *Physiol Plant* 128:8–17.
- Terán S., Rasmussen C., (1995) Genetic diversity and agricultural strategy in 16th century and present- day Yucatecan Milpa Agriculture. *Biodiv Conserv* 4:363–381.
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G., (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25, 4876-4882.
- Udoh P., Essien I. and Udoh F., (2005) Effects of *Carica papaya* (paw paw) seeds extract on the morphology of pituitary-gonadal axis of male Wistar rats, *Phytother Res*, 19(12), 1065-1068
- Verma R.J. and Chinoy N.J., (2001) Effect of papaya seed extract on microenvironment of cauda epididymis, *Asian J Androl*, 3(2), 143-146.
- Wai C.M., Ming R., Moore P.H., Paull R.E., Yu Q., (2010) Development of chromosome-specific cytogenetic markers and merging of linkage fragments in papaya. *Trop Plant Biol* 3:171-181.
- Wang J., Na J-K., Yu Q., Gschwend A.R., Han J., Zeng F., Aryal R., VanBuren R., Murray J.E., Zhang W., (2012) Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc Natl Acad Sci* 109: 13710–13715.

- Weston J., Ratle F. & Collobert R., (2008) Deep learning via semi-supervised embedding. Proceedings of the 25th international conference on Machine learning pp.1168–1175.
- Wuest S.E., Vijverberg k., Schmidt A. Weiss M., Gheyselinck J., Lohr M., Wellmer F., Rahnenführer J., Von Mering C., Grossniklaus U., (2010) Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. *Current Biology*, 20: 506-512.
- Yang Z., Rannala B., (2013) Improved Reversible Jump Algorithms for Bayesian Species Delimitation. *Genetics*, Vol. 194: 245-243.
- Yu Q., et al., (2008) Low X/Y divergence in four pairs of papaya sex-linked genes. *Plant J* 53:124-32.
- Zahn L. M., Leebens-Mack J., Depamphilis C.W., Ma H., Theissen G., (2005) To B or not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J. Hered.* 96, 225–240
- Zahn L.M., Leebens-Mack J.H., Arrington J.M., Hu Y., Landherr L.L., Depamphilis C.W., et al., (2006) Conservation and divergence in the AGAMOUS subfamily of MADS-box genes: evidence of independent sub- and neofunctionalization events. *Evol. Dev.* 8, 30–45.