

**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE - UFCG**  
**CENTRO DE EDUCAÇÃO E SAÚDE – CES**  
**UNIDADE ACADÊMICA DE BIOLOGIA E QUÍMICA - UABQ**

**ADEILMA FERNANDES DE SOUZA**

**SELEÇÃO *IN SILICO* DE GENES *MIRACULINA* DE *Citrus sinensis***  
**CANDIDATOS A ENSAIOS FUNCIONAIS**

**CUITÉ – PB**

**2015**

**ADEILMA FERNANDES DE SOUZA**

**SELEÇÃO *IN SILICO* DE GENES *MIRACULINA* DE *Citrus sinensis* CANDIDATOS A  
ENSAIOS FUNCIONAIS**

Monografia apresentada ao Curso de Licenciatura em Ciências Biológicas da Universidade Federal de Campina Grande, *Campus* Cuité, como forma de obtenção do grau de licenciada.

Orientadora: Profa. Dra. Magnólia de Araújo Campos

Coorientador: Dr. Marco Aurélio Takita

**CUITÉ – PB**

**2015**

FICHA CATALOGRÁFICA ELABORADA NA FONTE  
Responsabilidade Jesiel Ferreira Gomes – CRB 15 – 256

S729s Souza, Adeilma Fernandes de.

Seleção *in silico* de genes miraculina de *Citrus sinensis* candidatos a ensaios funcionais. / Adeilma Fernandes de Souza. – Cuité: CES, 2015.

57 fl.

Monografia (Curso de Licenciatura em Ciências Biológicas) – Centro de Educação e Saúde / UFCG, 2015.

Orientadora: Dra. Magnólia de Araújo Campos.  
Coorientador: Marco Aurélio Takita.

1. Genômica. 2. Banco de dados. 3. Mineração de dados. I. Título.

CDU 575

**ADEILMA FERNANDES DE SOUZA**

**SELEÇÃO *IN SILICO* DE GENES *MIRACULINA* DE *Citrus sinensis* CANDIDATOS A  
ENSAIOS FUNCIONAIS**

Monografia apresentada ao Curso de Licenciatura em Ciências Biológicas da Universidade  
Federal de Campina Grande, como forma de obtenção do grau de licenciada.

Aprovada em \_\_\_/\_\_\_/\_\_\_

**BANCA EXAMINADORA**

---

Prof<sup>ª</sup>. Dra. Magnólia de Araújo Campos (Orientadora)

---

Prof<sup>º</sup>. Dr. Rafael Trindade Maia

---

Prof<sup>º</sup>. Dr. Marcus José Conceição Lopes

Aos meus pais,  
José Batista Filho e Adeilza, pelo amor incondicional  
e por tudo que sempre fizeram e fazem para meu bem estar,  
ao meu vô Benedito Balbino (Nilton) por todo  
carinho, apoio e ensinamentos.

**DEDICO**

## AGRADECIMENTOS

Em primeira instância agradeço a Deus por estar sempre comigo me ajudando a seguir em frente, por não permitir que o desânimo se instale frente a meus dias, e por proporcionar a vivência de momentos inesquecíveis que serão lembrados com carinho em todo o percurso da minha vida.

A Universidade Federal de Campina Grande (UFCG-CES), Campus Cuité, por oportunizar a realização deste curso.

A minha orientadora Prof<sup>a</sup> Dr. Magnólia de Araújo Campos, por ter confiado em mim, por ser uma pessoa iluminada e ter me orientado não somente cientificamente, mais para vida através de suas palavras amigas e acolhedoras. O meu muito obrigada pela participação dessa fase tão marcante em minha vida.

Ao Dr. Marco Aurélio Takita, coorientador e pesquisador do Centro APTA de Citros Silvio Moreira, Cordeirópolis – São Paulo, pelo apoio e colaboração neste trabalho.

Aos professores membros da banca por terem aceitado participar da avaliação desse Trabalho de Conclusão de Curso (TCC)

A Erica Renata Nogueira de Sá, pelo apoio, incentivo e colaboração na realização deste trabalho.

A todos os professores que compõe o corpo docente do *campus* - Cuité que contribuíram efetivamente para minha formação.

Aos colegas do LBIotec, Thamara, Fernanda Furtado, Graciete Balbino, Ageu Santana, em especial a Rayane Abreu e Morgana Gomes, pela força e ajuda nesse trabalho, por fim a todos que compõe a equipe, obrigada pela convivência e incentivo.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela oportunidade de ingresso no Programa Institucional de Bolsa de Iniciação à Docência (PIBID).

A equipe PIBID da Escola André Vidal de Negreiros pela convivência e experiência adquirida no exercício prévio da docência, em especial a Márcio Frazão e Ângela Costa pelo auxílio e contribuição para meu crescimento enquanto bolsista.

Ao Bibliotecário do (CES), Jesiel Ferreira Gomes por dicas e ajuda relacionada a metodologia do trabalho científico, bem como a todos que compõe a equipe bibliotecária.

Aos funcionários que no anonimato são peças chave para o bom funcionamento do CES, a seu Vital pessoa por quem desenvolvi grande carinho.

A minha irmã de coração Thatiany Pereira, pelo amor e amizade, por todos os momentos inesquecíveis que vivemos que nos fez crescer como pessoas e futuras profissionais, pelo apoio nas horas difíceis

Aos meus grandes amigos (a) Aline Katiane, Noalixon Faustino, Serginho (Tilbert), por todos os momentos de aflição e diversão vividos, por estarmos juntos na alegria e na tristeza, no dinheiro e na sua ausência e em muitas loucuras da vida.

A Adriano Oliveira por representar uma fase boa da minha vida.

A toda minha família pelo carinho apoio e “puxões de orelha”, em especial a Tia Mariza por sempre me incentivar e acreditar no meu potencial e ter me dado forças em momentos que precisei.

A minha prima Jéssica Dayane, pelos conselhos, pela amizade, pela vivência juntas...

A meu vô.... Haa Vô como o senhor me ajudou, não apenas por ter me cedido sua casa, mais pelo carinho e cuidados quando precisei, quando estive doente, quando estava triste, quando fez a comida por que eu estava ocupada e não podia fazer, enfim por gestos e simples ações que me fazem ser eternamente grata ao senhor, obrigada... como carinhosamente chamo meu “Vozão”.

Aos meus pais e meus dois irmãos Alzeni Fernande e José Agnaldo (Gui) por serem meu apoio, minha base minha fortaleza e estarem comigo em todos os momentos, amo vocês!

Ao meu pequeno Anderson que a três anos veio para encher nossa casa de alegria, titia ama.

Enfim a todos que contribuíram de forma direta ou indiretamente para minha formação e crescimento pessoal!

*Quem pensa por si mesmo é livre,  
e ser livre é coisa muito séria,  
não se pode fechar os olhos,  
não se pode olhar para trás  
sem se aprender alguma coisa para o futuro.*

***Legião Urbana***



## RESUMO

O estudo da expressão de genes *Miraculina* em *Citrus sinensis* é de grande relevância, considerando-se a importância financeira atribuída às culturas de citros e variedades estudadas, bem como o envolvimento desses genes no desenvolvimento dos frutos e nas respostas de plantas a estresses bióticos e abióticos. O objetivo desse trabalho foi selecionar genes *Miraculina* de *Citrus sinensis* de interesse para estudos funcionais, usando análises *in silico* baseadas em informações dos bancos de dados CAP e CitEST. A estratégia empregada envolveu a identificação dos genes nos genomas por palavra-chave, análises de similaridade via BLASTx, análise de ORFs, análise de parâmetros físico-químicos e de domínios funcionais, análise da expressão em órgãos, bem como inferências filogenéticas a partir de alinhamento múltiplo de sequências de proteínas. Como resultados, no genoma da espécie vegetal *Citrus sinensis* foram encontrados 16 genes *Mir* identificados na cv. “Valência” e disponíveis no banco de dados CAP (*Citrus sinensis Annotation Project*). Dez genes *miraculina* da espécie *Citrus sinensis* do genótipo cultivar ‘Valência’ apresentaram sequências diferentes daquelas do genótipo cultivar ‘Pêra’ selecionadas no banco de dados CitEST. Dos 16 genes *Mir* de laranja doce ‘Valência’ identificados, quatro foram idênticos a quatro genes isolados da cultivar ‘Pêra’ e apresentam alto nível de expressão em frutos, folhas e flor sob variadas condições. Das quatro proteínas deduzidas de *C. sinensis* selecionadas e caracterizadas molecularmente, três apresentaram domínio para Inibidor de Tripsina da superfamília Kunitz. Das quatro proteínas deduzidas de *C. sinensis* selecionadas, três apresentaram pH teórico ácido e uma básico. Genes *CsMir1* e *CsMir2* representam bons candidatos para ensaios de atividade molecular *in silico*, por apresentarem melhores informações moleculares dentre os analisados viáveis ao seu funcionamento a nível de expressão, haja vista sua importância em aplicações biotecnológicas.

**Palavras-chave:** Banco de dados. Genômica. Mineração de dados.

## ABSTRACT

The study of miraculin gene expression in *Citrus sinensis* is of great importance, considering the financial importance of the citrus crop and studied varieties as well as the involvement of these genes in the development of the fruits and the responses of plants to biotic and abiotic stresses. The objective of this work was to select *Citrus sinensis* miraculin-like genes of interest for functional studies using *in silico* analysis based on information from databases CAP and CitEST. The strategy employed involved the identification of genes in keyword by genomes, analysis of similarity via BLASTx, ORFs analysis, analysis of physical-chemical parameters and functional domains, analysis of the expression in organs and phylogenetic inferences from Multiple alignment of protein sequences. As a result, the genome of the plant species *Citrus sinensis* found 16 genes Mir identified in cv. "Valencia" and available in CAP database (*Citrus sinensis Annotation Project*). Ten miraculin genes from the species *Citrus sinensis* genotype cultivar 'Valencia' presented different sequences from those of genotype cultivate Pera selected in CitEST database. Of the 16 genes sweet orange Mir 'Valencia' identified, four were identical to four single genes cultivar Pera and have a high level of expression in fruits, leaves and flower under varying conditions. Of the four proteins deduced from *C. sinensis* selected and characterized molecularly, three presented domain to Trypsin Inhibitor superfamily of Kunitz. Of the four proteins deduced from *C. sinensis* selected, three presented theoretical pI acid and a base. CsMir1 and CsMir2 genes represent good candidates for molecular activity assays *in silico*, by presenting best molecular information from the analyzed viable for it to function at the level of expression, given their importance in biotechnological applications.

**Keywords:** Data bank. Genomics. Data mining.

## LISTA DE FIGURAS

- Figura 1.** Sequencias de nucleotídeos e aminoácidos deduzida de um gene *Miraculina* de *Citrus sinensis* cv. ‘Valência’, selecionado a partir do banco de dados CAP, identificado como unigene Cs1g09420.1, representando uma ORF. Destaque em verde indica o códon de iniciação da tradução e em róseo indica o códon de parada da tradução. .... 34
- Figura 2.** Genes *Mir* de *C. sinensis* cv. ‘Valência’ com mais alto nível de expressão, de acordo com os dados do banco de dados CAP. .... 35
- Figura 3.** Expressão de 04 genes *Mir* em folha, fruto e flor de laranja doce cv. ‘Pêra’, equivalentes a 04 genes encontrados na cv. ‘Valência’, com base nas informações de ESTs presentes nos contigs disponíveis no banco de dados CitEST. .... 37
- Figura 4:** Alinhamento múltiplo das quatro sequencias de aminoácidos deduzidas de *C. sinensis* das cultivares “Valência” e “Pêra”. Em negrito, a sequencia N-terminal de peptídeo sinal, sombreado em amarelo indica a anotação para domínio inibidor de tripsina tipo Kunitz, sombreado em azul os pontos de fosforilação. Sublinhado indica sítios de N-glicosilação, sombreado em cinza indica as cisteínas conservadas, asteriscos e traços pretos indicam as cisteínas que formam pontes dissulfeto. .... 38
- Figura 5.** Dendrograma de proteínas miraculina de *C. sinensis* cv. ‘Valência’ com sequencias similares identificadas por BLASTx no GenBank (NCBI), identificadas na Tabela 2. .... 43

## LISTA DE TABELAS

- Figura 1.** Sequencias de nucleotídeos e aminoácidos deduzida de um gene *Miraculina* de *Citrus sinensis* cv. ‘Valência’, selecionado a partir do banco de dados CAP, identificado como unigene Cs1g09420.1, representando uma ORF. Destaque em verde indica o códon de iniciação da tradução e em róseo indica o códon de parada da tradução. .... 34
- Figura 2.** Genes *Mir* de *C. sinensis* cv. ‘Valência’ com mais alto nível de expressão, de acordo com os dados do banco de dados CAP. .... 35
- Figura 3.** Expressão de 04 genes *Mir* em folha, fruto e flor de laranja doce cv. ‘Pêra’, equivalentes a 04 genes encontrados na cv. ‘Valência’, com base nas informações de ESTs presentes nos contigs disponíveis no banco de dados CitEST. .... 37
- Figura 4:** Alinhamento múltiplo das quatro sequencias de aminoácidos deduzidas de *C. sinensis* das cultivares “Valência” e “Pêra”. Em negrito a sequencia N-terminal de peptídeo sinal, sombreado em amarelo indica a anotação para domínio inibidor de tripsina tipo Kunitz, sombreado em azul os pontos de fosforilação. Sublinhado indica sítios de N-glicosilação, sombreado em cinza indica as cisteínas conservadas, asteriscos e setas pretas indicam as cisteínas que formam pontes dissulfeto. .... 38
- Figura 5.** Dendrograma de proteínas miraculina de *C. sinensis* cv. ‘Valência’ com sequencias similares identificadas por BLASTx no GenBank (NCBI), identificadas na Tabela 2. .... 43
- Figura 6.** Dendrograma de proteínas miraculina de *C. sinensis* cv. ‘Pera’ com sequencias similares identificadas por BLASTx no GenBank (NCBI). .... 45

## LISTA DE ABREVIATURAS E SIGLAS

aa	Aminoácidos
APTA	Agência Paulista de Tecnologia dos agronegócios
BLAST	Do inglês, <i>Basic Alingnment Search Tool</i>
CAP3	Do inglês, <i>Sequence Assembly Program</i>
CAP	<i>Citrus sinensis</i> Annotation Project
cDNA	DNA complementar
CDS	Sequencias das regiões codificadas
CES	Centro de Educação e Saúde
CiTEST	Banco de dados de segmentos de sequencias expressas de citros
cv	Cultivar
DNA	Ácido dexorribonucleico
ESTs	Do inglês, <i>Expresses Sequence Tags</i> , segmentos de sequencias expressas
mRNA	RNA mensageiro
<i>Mir</i>	Gene <i>Miraculina</i>
Mir	Proteína <i>Miraculina</i>
NCBI	Do inglês, <i>National Center for Biotechnology Information</i>
ORFs	Do inglês, <i>Open Reading Frame</i>
RKPM	Do Inglês <i>Reds per Kb per million reads</i>
STI	Inibidores de proteinase Tripsina
UPGMA	Do Inglês <i>Unweighted Pair Group Method with Arithmetic mean</i>
UFMG	Universidade Federal de Campina Grande

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	15
1.1 Justificativa .....	15
<b>2 OBJETIVO GERAL</b> .....	17
2.1 Objetivos Específicos .....	17
<b>3 REVISÃO BIBLIOGRAFICA</b> .....	18
3.1 Proteínas do tipo Miraculinas.....	18
3.2 Família multigênica.....	19
3.3 Por que estudar proteínas miraculinas de Citros? .....	20
3.3.1 Importância dos citros para o Brasil e para a região Nordeste .....	22
3.4 Proteínas são produtos da expressão gênica.....	23
3.5 Importância da bioinformática para análises de sequencias gênicas.....	23
<b>4 MATERIAL E MÉTODOS</b> .....	26
4.1 Identificação e caracterização molecular de genes <i>Miraculina</i> de <i>C. sinensis</i> .....	26
4.2 Análise da expressão de genes <i>Mir</i> de <i>C. sinensis</i> .....	26
4.3 Análises de proteoma <i>in silico</i> para proteínas Mir preditas .....	27
4.3 Filogenia de proteínas Mir de <i>C. sinensis</i> .....	28
<b>5 RESULTADOS E DISCUSSÃO</b> .....	29
5.1 Identificação de genes miraculina de <i>Citrus sinensis</i> .....	29
5.2 Caracterização molecular de genes <i>miraculina</i> de <i>Citrus sinensis</i> .....	30
5.3 Estudo da expressão de genes <i>miraculina</i> de <i>Citrus sinensis</i> cultivares ‘Valência’ e ‘Pêra’ .....	34
5.4 Caracterização molecular de proteínas miraculinas de <i>Citrus sinensis</i> .....	38
5.5 Inferências filogenéticas de miraculinas de <i>Citrus sinensis</i> .....	42
<b>6 CONCLUSÃO</b> .....	47
<b>7 REFERÊNCIAS</b> .....	48

## 1 INTRODUÇÃO

Miraculina é uma proteína que tem a propriedade modificadora de sabor azedo em doce (KURIARA e TEERALSIP, 1988), que pode ser usada como adoçantes direcionados para diabéticos por adoçar diferencialmente e apresentar características superiores ao do mercado (GNANAVEL e PEDDHA, 2011). Além dessa propriedade, estudos comprovaram que membros da família de proteínas miraculinas possuem atividade de inibidor de tripsina do tipo Kunitz, atuando como agentes naturais de defesa da planta contra pragas e predadores, com potencial biopesticida para incorporação em culturas resistentes a pragas (GAHLOTH, 2011).

Estas proteínas apresentam relevante importância econômica é utilizada em diversas aplicações biotecnológicas, na indústria de processamento de alimentos como intensificadores de sabor, na alimentação de animais forrageiros, na indústria farmacêutica e na melhora o paladar de pacientes em tratamento de quimioterapia (SUN et al., 2006; BACHCHU et al., 2011; WILKEN e SATIROFF, 2012).

### 1.1 Justificativa

O gênero *Citrus* inclui algumas das mais importantes árvores frutíferas cultivadas em todo o mundo (CABALLERO, 2015). O Brasil é considerado mundialmente um dos maiores produtores de citros, a citricultura ocupa lugar de destaque no país em virtude da sua importância econômica em relação as exportações, e considerável geração de empregos (ROSSETTO, 2010). Maior parte da produção citrícola encontra-se concentrada nos Estados de Minas Gerais, Bahia, Sergipe, Rio Grande do Sul e São Paulo.

Dentre as espécies cítricas mais importantes comercialmente as cultivares de *Citrus sinensis*, laranja doce ‘Pêra’ e ‘Valência’ ganham considerável destaque. Pesquisas vêm sendo desenvolvidas com o intuito de encontrar soluções efetivas contra patógenos causadores de danos a essas culturas, nesse contexto, ferramentas de bioinformática e biologia molecular assumem ampla dimensão (ESPINDOLA, 2010).

A atualidade vive a “era pós genômica”, a ampla disponibilidade de informações geradas pelos sequenciamentos de espécies se torna instrumento de estudo de diversas áreas, como Agronomia, Medicina e Biologia Molecular com isso há grande necessidade de armazenamento e processamento adequado desses dados, com tudo a utilização desses bancos é crucial para tratamento dessas informações.

Com o grande volume de dados gerados, vem se tornando cada vez mais necessário a criação de bancos específicos. O Centro APTA de Citros Sylvio Moreira é referência nacional e internacional no setor agrícola de citros, mantém um dos maiores acervos de germoplasma de citros do mundo, contribui dando suporte à citricultura atual.

O banco de dados *Citrus sinensis* Annotation Project se encontra disponível online com dados do sequenciamento do genoma estrutural específico de Laranja – doce cultivar ‘Valência’, por ser um dos frutos mais importantes comercialmente, predominando na maioria dos países citrícolas. No entanto, fatores como restrições comerciais e problemas fitossanitários podem limitar os lucros e a expansão da citricultura (RANULFI, 2014). As pesquisas envolvendo a expressão de genes, sob diferentes condições e órgãos, buscam soluções para uma variedade de problemas principalmente de origem fitossanitária (CAICEDO, 2015).

Em virtude da pouca disponibilidade de fontes naturais desta proteína, pesquisadores vem buscando a proeza de produzi-la em outros organismos hospedeiros bem como *Escherichia coli* (KURIHARA, 1992), levedura, tabaco transgênico (KURIHARA E NIRASAWA, 1997) e morango (SUGAYA et al., 2008). Embora expressa nestes organismos, não manifesta a atividade de modificação de sabor.

O estudo da expressão do gene *Miraculina* em *Citrus sinensis* é de grande relevância, considerando a importância financeira atribuída as culturas de *citrus* e variedades estudadas, esse gene está envolvido na resposta da planta ao estresse biótico e abiótico, bem como o desenvolvimento dos frutos, e o nível de expressão sob determinadas condições refletem o estado fisiológico das células, tecidos ou órgãos.

Além disso o grande alvo dessa pesquisa é indicar genes *Mir* candidatos para ensaios funcionais *in silico*, por meio de modelagem e *docking* molecular com genes de interesse, visando suas aplicações biotecnológicas na cadeia produtiva dos citros, ou em outros setores como indústria alimentícia, entre outros



## 2 OBJETIVO GERAL

- Selecionar genes *Miraculina* de *Citrus sinensis* de interesse para estudos funcionais, usando análises *in silico* baseadas em informações dos bancos de dados CAP e CitEST.

### 2.1 Objetivos Específicos

- Identificar sequências de genes *Miraculina* de *C. sinensis* cultivares “Valência” e “Pêra”, a partir dos bancos de dados CAP (*Citrus sinensis* annotation Project) e CitEST (Citros ESTs), respectivamente, por palavras-chave;
- Caracterizar sequências de genes *Miraculina* de *C. sinensis* cultivares “Valência” e “Pêra” selecionadas, por similaridade via BLASTx e busca de ORFs;
- Caracterizar a expressão de genes *Mir* de *C. sinensis* cultivares “Valência” e “Pêra”, com base nas informações de bancos de dados CAP e CitEST;
- Realizar a caracterização molecular das proteínas Mir deduzidas a partir dos genes e selecionadas, por meio de análises proteômicas da predição de parâmetros físico-químicos, parâmetros de modificações pós-traducionais e parâmetros de domínios funcionais de família de proteínas.

### 3 REVISÃO BIBLIOGRAFICA

#### 3.1 Proteínas do tipo Miraculinas

Miraculina foi o nome dado a uma proteína isolada do fruto da planta milagre (*Synsepalum dulcificum*), atualmente classificada como *Richadella dulcifica*, descrita como uma glicoproteína, de peso molecular de aproximadamente 44 kD que tem a propriedade de modificar o sabor azedo em doce (KURIARA e BEILDLER, 1968)

Mais tarde, THEERASILP e KURIARA. (1988) descreveram a Miraculina (Mir) nativa como sendo uma proteína homodimérica constituída de 191 resíduos de aminoácidos com oligossacarídeos ligados a sítios por pontes dissulfetos com peso molecular de aproximadamente 24.6 kD, enquanto MASUDA et al. (1995) a descreveu com 220 aminoácidos. Sua sequência pura e desnaturada constitui-se de um tetrâmero, e um dímero, respectivamente, ambos apresentam atividade de modificação de sabor ácido, com máxima pH 3,0 e praticamente inativa ao pH 6,0 (IGETA et al., 1991; ANTONELA et al., 2010).

Os frutos de *R. dulcifica* foram tradicionalmente utilizados por povos indígenas para melhorar a palatatividade dos gêneros alimentícios ácidos e para adoçar bebidas ácidas (KURIHARA e BEIDLER, 1969; SUN et al., 2007). Atualmente no comércio japonês, a fruta milagrosa é encontrada *in natura* em pó e na forma de comprimidos. O pesquisador japonês hortícola, Mitsuharu Shimamura, produziu os comprimidos a partir da miraculina nativa para aumentar a disponibilidade desta proteína no mercado, encontrados sob a marca 'Miracle Fruit Tablet' (ONISHIKATSUMI-SHOKAI Corp. Osaka, Japão) (KURIHARA, 1992). Devido à presença de propriedades capazes de ocasionar efeitos adoçantes, com baixas concentrações de sacarose e de origem natural, esta glicoproteína pode ser utilizada em alimentos dietéticos e para diabéticos (FAUS, 2000; SUN et al., 2007).

Em virtude da pouca disponibilidade desta proteína por fontes naturais, pesquisas vêm sendo desenvolvidas a cerca da produção em larga escala por estratégias de engenharia genética utilizando *Escherichia coli* (KURIHARA, 1992), levedura, tabaco transgênico (KURIHARA e NIRASAWA, 1997), como organismos hospedeiros, embora a expressão dessa proteína por meio desses hospedeiros não apresenta atividade modificadora de sabor.

O acúmulo desta glicoproteína acontece de forma gradual durante o desenvolvimento do fruto atingindo alto nível na fase madura (HIRAI et al., 2011). A proteína Mir não é doce, porém uma vez que a língua humana é exposta a esta, sabores azedos possa a ser percebido

por fatores sensoriais como sabor doce, o efeito pode durar de 1 a 2 horas (BROUWER et al., 1968; KURIHARA et al., 1968).

Estudos moleculares estruturais e funcionais dessa glicoproteína mostram que resíduos de Histidina (His-30, His-60) são cruciais para ligações do receptor de sabor doce (T1R2, T1R3) nas papilas da língua preservando sua função modificadora de sabor, caso ocorra a substituição desses aminoácidos por alanina (Ala) em sua cadeia polipeptídica a função modificadora de sabor é perdida (KEISUKE ITO et al., 2007; PALADINO et al., 2008; ZE. H. et al., 2015).

Proteínas Mir apresentam importância relevante para aplicações biotecnológicas sendo utilizadas como adoçantes com características superiores aos do mercado, com direcionamento para diabéticos, visto que não seja necessário o uso de insulina, portando adoça diferencialmente (GNANAVEL e PEDDHA, 2011); intensificadores de sabor; suplementos de animais forrageiros na indústria de processamento de alimentos (BACHCHU et al., 2011; SUN et al., 2006), melhorando o paladar de pessoas em tratamento de quimioterapia (WILKEN e SATIROFF, 2012), além de apresentar atividade antifúngica (YE; WANG e NG, 2000).

As MirLP (*Miraculin-like proteins*) da superfamília de inibidores de proteinase tripsina tipo Kunitz são agentes naturais de defesa da planta contra pragas e predadores, sendo utilizadas como biopesticidas para a incorporação em culturas resistentes às pragas, alguns experimentos com *Murraya koenigii* mostra atividade inibitória do tipo tripsina contra duas pragas de insetos lepidópteros, *Helicoverpa armigera* e *Spodoptera litura* (GAHLOTH, 2011).

### **3.2 Família multigênica**

Genes miraculinas são encontrados nos genomas das espécies vegetais como famílias multigênicas. Uma família multigênica é formada por um grupo de genes descendentes de um mesmo gene ancestral, similares quanto a estrutura e funcionalidade (NEI e RONEY, 2005; DEL BEM, 2013). A presença de isoformas gênicas em genomas se deve a mutações do tipo substituição de nucleotídeos, duplicações e eventos de reparo, que podem acontecer ao logo do processo evolutivo (THOMPSON, 2006). O processo de duplicação gênica é fator 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 (KELLIS et al., 2004; DEL BEM, 2013).

De acordo com Ingram (1996), a pesquisa acerca do paradigma evolutivo das famílias multigênicas teve início com hemoglobinas e mioglobinas, estudos sugerem que os genes que codificam estas proteínas são filogeneticamente relacionados e divergiram conforme os genes foram duplicados adquirindo novas funções. Pesquisas mostram que as famílias multigênicas evoluíram por meio da Evolução Orquestrada (*concertd evolution*). Nesse modelo era necessário um mecanismo que homogeneizasse as diferentes cópias dentro de cada genoma, enquanto permitisse a divergência entre espécies ou população (NEY e ROONEY, 2005; DEL BEM, 2013).

O sequenciamento de genomas e a disponibilidade de ferramentas computacionais favorecem a compreensão acerca da evolução genética bem como permite abordagens comparativas sob a homologia dos genes (parálogos e ortólogos) e a constituição dessas famílias (YI LI, 2002; NEY e ROONEY, 2005). De acordo com Meyerowitz (2002), e Del Bem (2013), genes Ortólogos são definidos como aqueles que divergiram após evento de especiação e derivam de uma cópia ancestral comum e Parálogos como genes homólogos resultantes de um evento de duplicação gênica, restrito a apenas uma linhagem evolutiva.

Araújo (2010) constatou em seu trabalho realizado com *Arabidopsis thaliana*, planta modelo para estudos associados ao desenvolvimento do fruto, que a expressão de genes homólogos pertencentes a família multigênica MAD-box em frutos de *Citrus sinensis* variedade ‘Valência’. Conforme Oliveira (2011), MADS-box é uma família de genes que codifica fatores de transcrição que atuam como reguladores no desenvolvimento de vários organismos. Em eucariotos os genes desta família são conservados evolutivamente (PURUGGANAM et al., 1995).

### **3.3 Por que estudar proteínas miraculinas de Citros?**

Os citros (*Citrus* spp.) encontram-se agrupados na família *Rutaceae* que é composta de, aproximadamente, 155 gêneros e 1.600 espécies, amplamente distribuídas nas regiões tropicais, subtropicais e temperadas do mundo (NUNES, et al., 2005), subfamília Aurantioideae, tribo citreae e subtribo citrinae, com provável origem nas regiões do sudeste do continente asiático desde o centro da China ao Japão, estendendo-se até o leste da Índia, Nova Guiné, Austrália e Àsia tropical (CAVALCANTI, 2005; MAGALHÃES, 2009). Conforme escritos encontrados na China essas plantas provavelmente foram levadas para a Europa na época das cruzadas e chegaram ao Brasil trazidos pelos portugueses no século XVI (MATOS JUNIOR et al., 2005; LIMA, 2013).

As espécies do gênero *Citrus*, *Poncirus* e *Fortunella* reproduzem-se sexuadamente por autopolinização e polinização cruzada, assexuadamente por apomixia nucelar adventícia e agronomicamente por propagação vegetativa. Suas sementes possuem tanto embriões zigóticos, como apomíticos, apresentando, em geral, apomixia facultativa, com número variável de embriões entre um e doze (OLIVEIRA, 2006).

A filogenia e a taxonomia dos *Citrus* são bastante complexas, em função da alta variabilidade do gênero, decorrente de hibridações naturais interespecíficas, mutações espontâneas, alterações cromossômicas estruturais, apomixia e cultivo em ampla área geográfica (OLIVEIRA e RADMANN, 2005). Há divergências quanto a classificação taxonômica dos *citrus*, segundo Engler (1931), esse gênero constitui-se por 11 espécies, (SWINGLE e REECE, 1967) subdividiu em dois subgêneros *Papeda* e *Citrus* com 16 espécies, contudo este é o sistema de classificação mais adotado, Tanaka (1977), propôs a existência de 159 espécies para os citros e três para Barret e Rhodes (1976), (BASTINAEL et al., 2001; SORIANO, 2010; LIMA, 2013).

Atualmente *C. medica*, *C. reticulata* e *C. grandis* têm sido aceitos como subgêneros dos *Citrus* (MACHADO et al., 2005). Os demais táxons desses subgêneros são híbridos naturais entre tais espécies ou entre representantes sexualmente compatíveis de outros subgêneros (OLIVEIRA, 2006; FARIA, 2007). Os citros de maior interesse comercial são formados por um amplo grupo de espécies de híbridos principalmente dos gêneros *Poncirus*, *Citrus* e *Fortunella*, sendo as espécies do gênero *Citrus* as de maior importância econômica, as quais podem ser destacadas: as laranjas (*Citrus sinensis*), as tangerinas (*Citrus reticulata* e *Citrus deliciosa*), os limões (*Citrus limon*), as limas ácidas como Tahiti (*Citrus latifolia*) e galego (*Citrus aurantifolia*) e doces como a lima da pérsia (*Citrus limettioides*), o pomelo (*Citrus paradisi*), a cidra (*Citrus medica*), a laranja azeda (*Citrus aurantium*) e as toranjas (*Citrus grandis*) (MAGALHÃES, 2009).

As laranjas doces (*Citrus sinensis* (L.) Osbeck) são consideradas comercialmente uma das mais importantes frutas cítricas, predominando na maioria dos países citrícolas com, aproximadamente, dois terços dos plantios. Os principais cultivares destinados a produção de suco no Brasil é basicamente, representadas pelas variedades de laranja doce ‘Pêra’, ‘Valência’, ‘Natal’ e ‘Hamilin’ (PIO et al., 2005; FÁVERO, 2010).

A citricultura é uma das principais atividades de destaque na economia brasileira, no entanto problemas de ordem fitossanitária, além de restrições comerciais podem limitar a produção e exportação (RANULFI, 2014). Dentre as doenças que acometem as plantas cítricas considera-se de bastante relevância a Morte Súbita, Greening, Clorose Variegada dos

Citros e o Cancro Cítrico *Huanglongbing* entre outras, vem causando significantes perdas e em alguns casos levando a improdutividade (NEVES et al., 2010; OLIVEIRA, 2014; CAICEDO, 2015).

O genoma estrutural e funcional de citros de várias espécies vem sendo sequenciado visando compreender o funcional e genética dos citros na busca de soluções para citricultura mundial.

### **3.3.1 Importância dos citros para o Brasil e para a região Nordeste**

A indústria citrícola é uma das principais atividades do agronegócio Brasileiro, sendo o suco concentrado congelado sua principal “*commodity*” de exportação (FARIA, 2007). Dentre as regiões produtoras, a região sudeste tem maior destaque, com ênfase no Estado de São Paulo, que é responsável por aproximadamente 80% da produção, a região Nordeste, com a Bahia e Sergipe, juntas corresponde a 8,65% da produção nacional. Em virtude da ausência de pragas de grande significância como as encontradas na região Sudeste considerado maior centro produtor, a citricultura no Nordeste apresenta ampla potencialidade de crescimento (AZEVEDO, 2007).

De acordo com Alvino (2011), a região da Paraíba está em 8º lugar dentre os estados produtores do Nordeste, tendo como foco à produção de tangerinas, porém, a produtividade ainda é baixa, devido a fatores ambientais como o déficit hídrico que ocorre durante mais de seis meses no ano, coincidindo, geralmente, com temperaturas elevadas, necessitando de irrigação para se obter maiores níveis de produtividade. Problemas como crescimento de anomalias fisiológicas nas plantas cítricas, redução do crescimento, desenvolvimento e produtividade pode ser ocasionado pela qualidade da água utilizada na irrigação por apresentar uma elevada concentração de sais, além de problemas fitossanitários que ocorrem em pomares por todo o Brasil.

A produção de Citros na Paraíba ainda é limitada, devido principalmente a fatores ambientais como déficit hídrico e altas temperaturas com ciclos que duram mais de seis meses, assim o estudo da expressão de genes *Mir* pode ser uma alternativa de encontrar melhorias para problemas como este (SUN et al., 2006; PALADINO et al., 2007).

### **3.4 Proteínas são produtos da expressão gênica**

A Expressão Gênica ocorre quando a informação contida em um gene (sequências de nucleotídeos do DNA) é processada em um produto gênico, bem como uma proteína. Desta maneira a mesma apresenta duas distintas etapas: a produção do mRNA (transcrição) e a síntese da proteína (tradução). Em eucariotos, o processo acontece do núcleo para o citoplasma, através das etapas de transcrição onde a informação genética é copiada utilizando um filamento molde da dupla hélice do DNA, o RNA mensageiro é sintetizado de modo complementar ao filamento molde. Após os eventos transcripcionais e a produção da molécula que atuará como mensageira (mRNA) levando a informação genética ao citoplasma onde ocorrerá os demais estágios da síntese proteica (tradução) gerando a proteína como produto final de interesse para a célula (BORGES, 2006).

A análise da expressão genica é de grande interesse para diversas áreas científicas, através dela pode-se compreender as várias formas de polimorfismos que podem ser de origem deletéria, mutações, substituição de bases, bem como varrições no número de sequencias repetidas (METZGER et al., 2006).

Muitas tecnologias vêm sendo desenvolvidas, dentre estas a técnica de *Microarrays*, ou *chips* de DNA ganham destaque. Através de novas metodologias de investigação pesquisadores podem examinar comparativamente a expressão de milhares de genes de forma simultânea a nível global em células de vários organismos, sob diferentes condições e órgãos (GUINDALINI e TUFIK 2007).

### **3.5 Importância da bioinformática para análises de sequencias gênicas**

A bioinformática foi definida por Prodoscimi et al. (2003) e Alves (2013) como uma nova ciência que engloba a união de diversas linhas do conhecimento, incluindo a matemática, a estatística a engenharia de softwares, a biologia molecular e a ciência da computação. Essa nova ciência possibilita a demonstração das funções gênicas e suas relações com proteínas, através da análise computacional de sequencias de DNA, RNA e proteínas (ALVES, 2013).

Junto com essa nova ciência veio a necessidade de armazenamento e análise dos dados gerados a partir de seus estudos e aplicações em diversas áreas do conhecimento humano medicina, agronomia, biologia molecular e afins (ARAÚJO, 2004). De acordo com Mello et

al. (2008), essa área do conhecimento passou a ter destaque na década de 90 com o aparecimento do Projeto Genoma. Lifschitz (2006) propôs que o objetivo central da bioinformática é a análise de dados biológicos por meio de técnicas avançadas da computação e estudos científicos envolvendo genômica, transcriptoma, e proteômica. O sequenciamento de nucleotídeos é fator relevante para o tratamento computacional pós-genômico, interessante ao desenvolvimento da Biologia Molecular (SILVEIRA e LIBRELOTTO, 2007).

Na atualidade há uma vasta quantidade de informações disponíveis de sequências de aminoácidos e nucleotídeos resultantes de projetos genomas, constituindo uma elevada quantidade de informações a cerca dos mais diversos organismos. Esta vasta quantidade de informações são armazenadas em bancos de dados disponíveis nacional e internacionalmente de forma atualizada e de fácil acesso (PROSDOCIMI et al., 2003).

Os bancos de dados podem ser classificados em primários aqueles em que são depositados sequências de nucleotídeos, aminoácidos ou proteínas sem nenhum processamento a exemplo o GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) e secundários formados a partir de bancos primários SWISS-PROT (PROSDOCIMI et al., 2003). Uma das ferramentas de busca de dados amplamente usada é o sistema de *Information Retrieval* do Entrez PubMed fornecido pelo NCBI (*National Center For Biotechnology Information*) (<http://www.ncbi.nlm.nih.gov/>), bem como o google scholar (<https://scholar.google.com.br/>) específico para busca de literatura acadêmica (ESPINDOLA et al., 2010).

Os crescentes investimentos na manutenção e construção de bancos de dados públicos faz-se necessário devido à grande variabilidade de informações geradas propiciando a formação de bancos mais específicos, como o CiTEST e o *Citrus sinensis* annotation Project (CAP).

CiTEST é um grande projeto de sequenciamento do transcriptoma, que objetiva integrar conhecimentos sobre mapeamento genético com novos dados sobre genômica funcional e comparativa de espécies de Citros e gêneros afins e a resposta dessas plantas sob estresses biótico e abióticos. O mesmo inclui o sequenciamento completo do genoma de agentes patogênicos (vírus e bactérias) de citros para estudo de interações planta-patógeno (REIS, 2007). Já o CAP é um banco de dados considerado novo construído em 2012 após a conclusão do sequenciamento do genoma duplo haplóide de laranja doce, fornece uma plataforma integrada com a organização de dados genômicos, incluindo visão geral dos cromossomos. Tem o intuito de proporcionar à comunidade científica uma anotação precisa da sequência do genoma dessa espécie (cultivar ‘Valência’) (WANG et al., 2014).



A bioinformática além de oferecer informações para armazenar elevados volumes de dados tem a disponibilidade ferramentas de diversas análises de sequencias gênicas, permitindo o alinhamento e reconhecimento das funções dos genes bem como a evidência de fatores evolutivos.

## 4 MATERIAL E MÉTODOS

As análises computacionais das sequências gênicas foram realizadas no Laboratório de Biotecnologia – LBiotec, do Centro de Educação e Saúde da Universidade Federal de Campina Grande, em Cuité PB, usando computador com sistema operacional Windows 7 Professional, variante 64 bits, conectado a rede de internet.

### 4.1 Identificação e caracterização molecular de genes *Miraculina* de *C. sinensis*

As sequências gênicas estudadas foram obtidas a partir de dois bancos de dados: 1) o do projeto genoma duplo haploide de laranja-doce, *Citrus sinensis* Annotation Project (CAP Database), da China, disponível em <http://citrus.hzau.edu.cn/orange/>; e 2) e do genoma funcional e comparativo dos citros (CitEST Database), realizado pelo Centro APTA de Citros Sylvio Moreira, Cordeirópolis SP, Brasil, acessível mediante senha autorizada por colaboração em pesquisas em <http://biotecnologia.centrodecitricultura.br/>. A busca por genes *Miraculina* (*Mir*) nos dois bancos de dados foi feita usando a palavra-chave “*miraculin*”. As sequências gênicas obtidas foram coletadas da forma como se encontravam acessíveis nos referidos bancos, arquivo texto no formato FASTA, para posteriores análises individuais.

Para a caracterização molecular dos genes *Mir*, as sequências de interesse para este estudo foram submetidas a uma análise de similaridade por comparação via BLASTx (*Basic Local Alignment Search Tools*) (ALSTUHL, 1990), disponível em <http://blast.ncbi.nlm.nih.gov/blast/Blast.cgi> com a finalidade de encontrar sequências homólogas a miraculina no banco de dados do GenBank do NCBI (*National Center for Biotechnology Information*). Em seguida, todas as sequências dos unigenes de ‘Valência’ e de ‘Pêra’ foram analisados quanto à presença de ORFs (*Open Reading Frame*) completas, usando o Programa ORFFinder, acessível em <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>, e as sequências de aminoácidos preditas a partir das sequências dos unigenes foram identificadas para posteriores análises proteômicas *in silico*.

### 4.2 Análise da expressão de genes *Mir* de *C. sinensis*

O estudo da expressão de genes *Mir* em *C. sinensis* foi realizado com base nas informações disponibilizadas nos bancos de dados CAP e CitEST para as cultivares

‘Valência’ e ‘Pêra’, respectivamente. Os dados de expressão de genes *Mir* da laranja doce ‘Valência’ foram obtidos na forma de gráficos com valores de RKPM (*Reads per Kb per Million reads*) relativos aos valores normalizados da expressão baseada no número de *reads* obtidas a partir de bibliotecas de RNASeq de calo, fruto, folha e flor. Neste caso, as informações computacionais foram validadas por vários experimentos laboratoriais (XU et al., 2013; WANG et al. 2014). Já as informações de expressão gênica da laranja doce ‘Pêra’, a estimativa da expressão gênica foi dada pelo número de ESTs (*Expressed Sequence Tags*), que formaram os unigenes *Mir*, isolados de diferentes bibliotecas de cDNA construídas de vários órgãos (folha, flor, frutos caule e raiz) sob diferentes condições de estresses biótico e abiótico (TARGON et al., 2007) e os dados analisados e apresentados como gráficos gerados no excel 2013.

#### 4.3 Análises de proteoma *in silico* para proteínas *Mir* preditas

A caracterização molecular de sequencias de aminoácidos deduzidas via ORFFinder (item 4.1) foi realizada por predição de parâmetros físico-químicos, parâmetros de modificações pós-traducionais e parâmetros de domínios funcionais de família de proteínas. Os parâmetros físico-químicos analisados foram número de aminoácidos, massa molecular, pH teórico e meia vida, usando a ferramenta ProtParam (GASTEIGER, 2005), disponível em <http://web.expasy.org/protparam/>. Os parâmetros de modificações pós-traducionais analisados foram busca e localização por 1) sítios de clivagem de peptídeos líder, usando o Programa SignalP versão 3.0 (BENDTSEN, 2004), disponível em <http://www.cbs.dtu.dk/services/SignalP-3.0/>; 2) sítios de N-glicosilação, usando o Programa NetNGlyc 1.0 server (GUPTA, 2004), disponível em <http://www.cbs.dtu.dk/services/NetNGlyc/>; 3) sítios de fosforilação, usando a ferramenta NetPhos 2.0 server (BLOM, 1999), disponível em; <http://www.cbs.dtu.dk/services/NetPhos/> e 4) sites de miristoilação, usando o software Myrstoilator (BOLOGNA, 2004), disponível em <http://web.expasy.org/myrstoilator/>. Para a análise de domínios funcionais de famílias de proteínas utilizou-se o Programa ScanProsite (SIGRIST, 2012), disponível em <http://prosite.expasy.org/prosite.htm/>.

Para uma melhor visualização e comparação dos resultados de proteoma, um alinhamento múltiplo de sequencias foi realizado para sequencias de aminoácidos *Mir* preditas, usando o Programa Clustalw (*Multiple Sequence Alignment*) (THOMPSON, 1994),

disponível em <http://www.genome.jp/tools/clustalw/>, e os dados de proteoma foram indicados para cada sequência analisada. A partir do alinhamento múltiplo de sequências é possível evidenciar a conservação de regiões características da família da proteína em estudo.

#### **4.3 Filogenia de proteínas Mir de *C. sinensis***

Com o intuito de obter informações filogenéticas, um alinhamento múltiplo foi realizado entre as proteínas Mir deduzidas de *C. sinensis* cv. 'Valência' e as sequências de melhor similaridade via BLASTx disponíveis no GenBank, utilizando o Clustalw. Os dados do alinhamento foram utilizados para gerar uma árvore filogenética (dendrograma de similaridade) enraizada com comprimento de braço pelo método de UPGMA (de *Unweighted Pair Group Method with Arithmetic mean*), usando os critérios de agrupamento hierárquico do vizinho mais próximo, vizinho mais distante da média aritmética entre grupos de pares não ponderados (THOMPSON, 1994), disponível em <http://www.genome.jp/tools/clustalw/>.

## 5 RESULTADOS E DISCUSSÃO

Este trabalho descreve uma estratégia *in silico* para a seleção e indicação de genes novos *Miraculina* de *Citrus sinensis* das cultivares ‘Valência’ e ‘Pêra’, por meio da identificação, caracterização molecular e análise da expressão *in silico*, discutindo de maneira comparativa os dados disponíveis nos bancos de dados CAP (*Citrus sinensis Annotation Project*), Chinês, e CitEST (*Citrus Expressed Sequence Tags*), brasileiro, respectivamente.

### 5.1 Identificação de genes miraculina de *Citrus sinensis*

A estratégia de busca por palavra-chave com o termo “miraculin” no banco de dados chinês levou a descoberta de 16 unigenes *miraculina* (*Mir*) diferentes, sendo 11 contigs e 5 singlets (Tabela 1). Entretanto, usando esta mesma estratégia, RODRIGUES (2012) identificou 25 unigenes no banco de dados brasileiro do genoma funcional dos citros (CitEST), sendo 14 contigs e 11 singlets.

**Tabela 1:** Identificação dos unigenes *miraculina* de *Citrus sinensis* isolados das cultivares ‘Valência’ e ‘Pêra’ (RODRIGUES, 2012), de acordo com os bancos de dados CAP e CitEST, respectivamente.

Banco de dados	Cultivar	Identificação dos Unigenes			
		Contigs		Singlets	
CAP	Valência	Cs1g09420.1	Cs5g13960.1	orange1.1t04805	
		Cs1g09420.2	Cs5g16780.1	orange1.1t02835	
		Cs5g13930.1	Cs5g13890.1	orange1.1t04806	
		Cs2g16250.1	Cs5g16960.1	orange1.1t02833	
		Cs5g16770.1	Cs5g16920.1	orange1.1t01406	
		Cs5g16850.1			
CitEST	Pêra	CAS-Cs-107510.1	CAS-Cs-107297.1	CAS-Cs-131323.1	CAS-Cs-122084.1
		CAS-Cs-108140.1	CAS-Cs-106143.1	CAS-Cs-120059.1	CAS-Cs-124320.1
		CAS-Cs-103575.1	CAS-Cs-107757.1	CAS-Cs-124740.1	CAS-Cs-119802.1
		CAS-Cs-105567.1	CAS-Cs-106095.1	CAS-Cs-124786.1	CAS-Cs-125583.1
		CAS-Cs-106984.1	CAS-Cs-107308.1	CAS-Cs-124869.1	CAS-Cs-131338.1
		CAS-Cs-107226.1	CAS-Cs-107556.1	CAS-Cs-119964.1	
		CAS-Cs-103444.1	CAS-Cs-109117.1		

Para compreender a observação de número de unigenes diferentes encontrado nos dois genótipos da mesma espécie, cultivares ‘Valência’ e ‘Pêra’, várias análises foram feitas usando somente as sequencias das regiões codificadoras (CDS), na tentativa de encontrar os pares de genes idênticos de cada genótipo, incluindo: 1) alinhamentos múltiplos de todas as sequencias de nucleotídeos, obtidas nos dois bancos de dados, e de aminoácidos deduzidos, seguidos por agrupamentos de ambas mostrados em dendrogramas de similaridades; 2) alinhamentos dois a dois usando todas as sequencias de nucleotídeos, obtidas nos dois bancos de dados, e de aminoácidos deduzidos; e 3) clusterização de todas as sequencias de nucleotídeos usando o programa CAP3 (*Contig Assembly Program*), disponível em <http://doua.prabi.fr/software/cap3>. Como resultado, somente foram encontrados unigenes idênticos para 6 genes, sendo os demais diferentes (dados não mostrados). Portanto, a diferença observada no número de genes *Mir* de laranja doce ‘Valência’ e de ‘Pêra’ ocorreu também ao nível de sequencias de nucleotídeos e de aminoácidos deduzidos para a maioria deles.

Uma vez que no banco de dados do CAP possui informações de mapeamento dos genes nos cromossomos, os dados são confiáveis e, portanto, é correto considerar que miraculinas são codificadas por uma família multigênica no genoma de *C. sinensis*, composta por 16 genes diferentes no genótipo da cultivar ‘Valência’ (WANG et al., 2014). Após sequenciamento, as *reads* são submetidas a uma clusterização, para gerar os grupos unigenes, que podem ser formados por mais de uma sequencia considerada igual (contigs) ou por apenas uma sequencia (singlet). Assim, a baixa qualidade de algumas sequencias de DNA leva estas a não se agruparem em *Contigs* de maneira correta, o que pode explicar, em parte, o acontecido com as sequencias da cv. ‘Pêra’ do banco de dados CitEST.

## 5.2 Caracterização molecular de genes *miraculina* de *Citrus sinensis*

As sequencias de cada um dos unigenes encontrados em ambos os bancos de dados foram submetidas a análises de BLASTx e as sequencias de aminoácidos deduzidas compartilharam elevada similaridade com outras miraculinas depositadas no GenBank (NCBI), de modo a atualizar informações dos bancos de dados CAP e CitEST. Na Tabela 2 encontra-se demonstrada as melhores similaridades via BLASTx de miraculinas de *C. sinensis* da cultivar ‘Valência’.

**Tabela 2:** Similaridade via BLASTx de miraculinas de *Citrus sinensis* cultivar ‘Valência’ selecionadas no banco de dados CAP.

UNIGENE	COMPRI-MENTO DA READ (pb)	NOME DO GENE	ORGANISMO	NÚMERO DE ACESSO	SIMILARIDADE/BLASTx
Cs1g09420.1	1159	hypothetical protein CICLE_v10002395mg	<i>Citrus clementine</i>	XP_006432185.1	205/205(100%)
Cs1g09420.2	1016	hypothetical protein CICLE_v10002395mg	<i>Citrus clementine</i>	XP_006432185.1	224/224(100%)
Cs2g16250.1	836	Trypsin inhibitor	<i>Murraya koenigii</i>	ACS37303.1	132/158(84%)
Cs5g13890.1	1244	Hypothetical protein CICLE_v10022211mg	<i>Citrus clementine</i>	XP_006442410.1	210/213(99%)
Cs5g13930.1	652	Hypothetical protein CICLE_v10022211mg	<i>Citrus clementine</i>	XP_006442410.1	177/213(83%)
Cs5g13960.1	615	PREDICTED: 21 kDa seed protein-like	<i>Citrus sinensis</i> (cv. Valência)	XP_006477857.1	165/179(92%)
Cs5g16770.1	1296	Miraculin-like protein 2	<i>Citrus x paradise</i>	AAG38518.1	217/218(99%)
Cs5g16780.1	1133	Miraculin-like protein	<i>Citrus tamurana</i>	BAQ21278.1	228/232(98%)
Cs5g16850.1	1328	Miraculin-like protein 2	<i>Citrus japonica</i>	AFR54471.1	219/223(98%)
Cs5g16920.1	921	Miraculin-like protein 2	<i>Citrus japonica</i>	AFR54471.1	195/224(87%)
Cs5g16960.1	837	Hypothetical protein CICLE_v10024002mg	<i>Citrus clementine</i>	XP_006441469.1	198/221(90%)
orange1.1t01406	3234	Hypothetical protein VITISV_008177	<i>Vitis vinifera</i>	CAN63013.1	224/367(61%)
orange1.1t02833	787	Miraculin-like	<i>Citrus sinensis</i> (cv. Valência)	XP_00649506.1	195/198(98%)
orange1.1t02835	771	Hypothetical protein CICLE_v10023257mg	<i>Citrus clementine</i>	XP_006441435.1	171/174(98%)
orange1.1t04805	846	Hypothetical protein CICLE_v10022284mg	<i>Citrus clementine</i>	XP_006441459.1	200/206(97%)
orange1.1t04806	830	PREDICTED: Miraculin-like	<i>Citrus sinensis</i> (cv. Valência)	XP_006493430.1	195/198(98%)

Como é possível observar, as miraculinas de laranja doce cv. ‘Valência’ apresentaram elevada similaridade (61 a 100%) com outras miraculinas da superfamília de inibidores de proteinase Tripsina do tipo Kunitz de *Citrus clementine* (07 sequencias), *Citrus sinensis* cv. ‘Valência’ (3), *Citrus japonica* (2), Citrus x Paradise (1), *Citrus tamurana* (1), *Vitis vinifera* (1) e *Murraya koenigii* (1).

A análise de BLASTx para sequencias *Mir* de laranja doce ‘Pêra’ foi atualizada levando a confirmação de que as 25 sequencias apresentaram similaridade com proteínas *Mir*

de outras espécies de citros, sendo somente uma (01) com *Murraya koenigii*. Entretanto, uma análise mais minuciosa revelou que nove (09) sequências encontradas por palavra-chave Miraculin no banco de dados CitEST apresentaram similares apenas de trecho curtos, variando de 38 a 134 pb, as quais foram eliminadas das análises. Portanto, 16 unigenes *Mir* também foram encontrados em laranja doce ‘Pêra’ compartilhando elevada similaridade (81 a 100%) com outras proteínas do tipo miraculinas da superfamília de inibidores de proteinase Tripsina do tipo Kunitz de *Citrus clementine* (08 sequências), (3), *Citrus japonica* (2), *Citrus x Paradise* (2), *Citrus tamurana* (2) e *Citrus jambiri* (1) (Dados completos não apresentados).

Assim, de acordo com a análise de BLASTx, a diferença observada entre genes *Mir* de laranja doce ‘Valência’ e de ‘Pêra’ anteriormente ao nível de alinhamentos de sequências de nucleotídeos e de aminoácidos deduzidos para a maioria deles, ocorreu também ao nível de similaridade com sequências de outras proteínas do tipo *Mir* disponíveis no GenBank.

Uma explicação para esta observação poderia ser a existência de variabilidade intraespecífica que ocorre em cultivares dessa espécie por mutação espontânea, gerando novos genótipos ou cultivares, devido ao fato de que *C. sinensis* não ser uma espécie pura (PEREIRA, 2012). Além disso, de acordo com Xu et al. (2013), a cultivar Valência possui elevado grau de heterozigose como resultado da origem da laranja doce ser por hibridização interespecífica entre toranja (*C. maxima*) e tangerina Ponkan (*C. reticulata*). Por outro lado, apesar da citada origem de *C. sinensis*, em ambos os casos observou-se, ainda, a elevada similaridade entre sequências de aminoácidos dos dois genótipos estudados com a espécie *C. clementina*, ao invés de ser com as espécies parentais.

Em seguida, a análise de cada uma das sequências gênicas quanto à presença de uma ORF (*Open Reading Frame*) completa revelou que todos os unigenes *Mir* de *C. sinensis* cv. ‘Valência’, identificados no banco de dados CAP, apresentaram sequências abertas de leitura para a síntese de proteínas. Similarmente, 15 unigenes da cultivar ‘Pêra’ identificados no CitEST apresentaram ORFs completas, de um total de 16 unigenes selecionados via BLASTx, sendo que somente um deles apresentou ORF parcial.

De modo geral, o comprimento das ORFs gênicas *Mir* de laranja doce ‘Valência’ variou de 479 a 3234 pb, as quais possivelmente codificam proteínas contendo de 159 a 1077 aminoácidos deduzidos, enquanto que o comprimento de ORFs completas de ‘Pêra’ variou de 549 a 711 pb, as quais codificam proteínas preditas contendo de 182 a 236 aminoácidos (Tabela 3).



**Tabela 3:** Comprimento em pares de bases (pb) das ORFs gênicas e número de aminoácidos (aa) das proteínas preditas a partir de genes *Mir* de *Citrus sinensis* isolados das cultivares ‘Valência’ e ‘Pêra’ e identificados a partir dos bancos de dados CAP e CitEST.

Genes <i>Mir</i> isolados da Cultivar Valência			Genes <i>Mir</i> isolados da Cultivar Pêra		
Unigene	ORF (pb)	Proteína (aa)	Unigene	ORF (pb)	Proteína (aa)
Cs1g09420.1	618	205	CAS- CS-106095.1	618	205
Cs1g09420.2	675	224	CAS-Cs-107297.1	672	223
Cs2g16250.1	479	159	CAS-Cs-103444.1	636	211
Cs5g13890.1	642	213	CAS-Cs-106143.1	672	223
Cs5g13930.1	642	213	CAS-Cs-107226.1	642	213
Cs5g13960.1	615	204	CAS-Cs-106984.1	549	182
Cs5g16770.1	711	236	CAS- CS-103575.1	711	236
Cs5g16780.1	699	232	CAS-CS-107308.1	699	232
Cs5g16850.1	672	223	CAS-CS-107757.1	672	223
Cs5g16920.1	672	223	CAS-Cs-107510.1	618	205
Cs5g16960.1	837	278	CAS-Cs-107556.1	699	232
orange1.1t01406	3234	1077	CAS-Cs-108140.1	672	223
orange1.1t02833	621	206	CAS-Cs-119964.1	662	216
orange1.1t02835	633	210	CAS-Cs-119802.1	579	192
orange1.1t04805	621	206	CAS-Cs-125583.1	648	215
orange1.1t04806	621	206			

Como demonstrado na Tabela 3, um unigene da cultivar ‘Valência’ do tipo *singlet*, identificado como orange1.1t01406, apresentou comprimento muito mais elevado do que a média. Esta sequência compartilhou a menor similaridade via BLASTx (61%, 224/367) com uma Hypothetical protein VITISV\_008177 de *Vitis vinifera*.

Na Figura 1 está representada a ORF de um gene *Mir* isolado de *C. sinensis* cv. ‘Valência’ (unigene Cs1g09420.1), de comprimento de 618 pares de bases (pb), que codifica uma proteína contendo 205 aminoácidos deduzidos. Nenhum dos genes *Mir* de ambas as cultivares não apresentou introns.

```

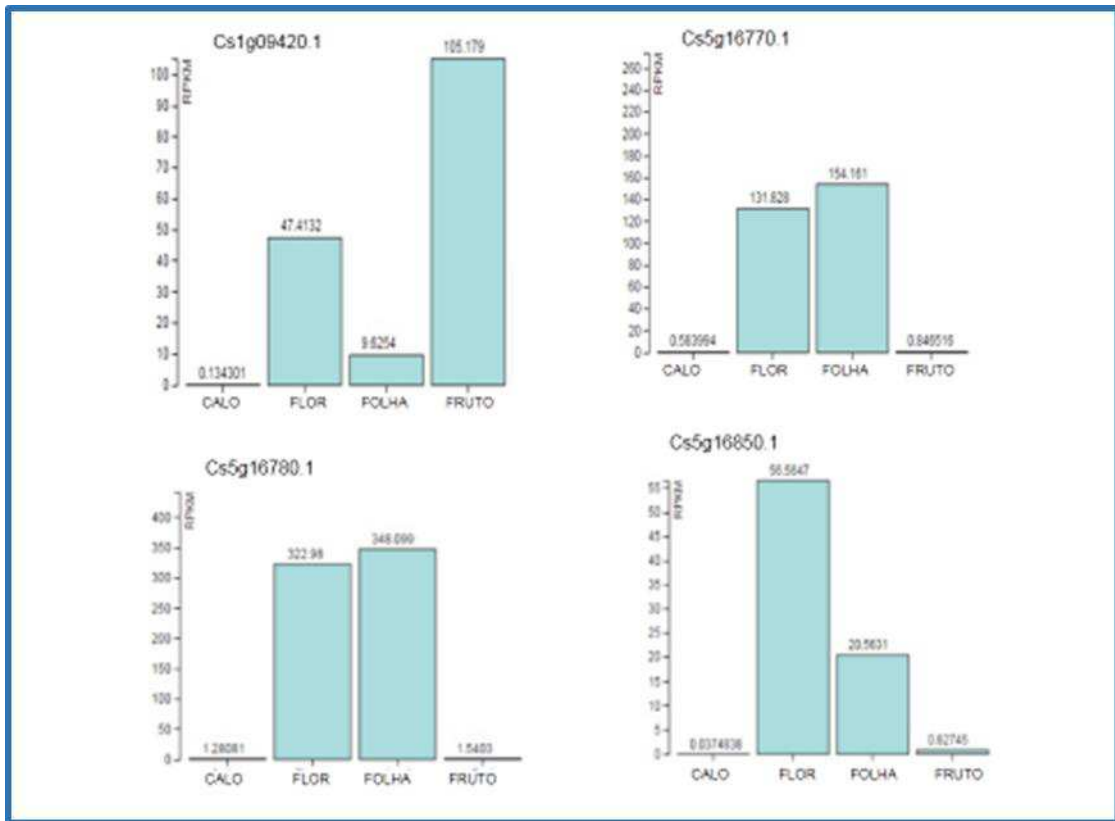
1 atgaagacttcattcgtagtaaaactctcattccttataccttgcc
M K T S F V V K L S F L I L A
46 ttggccacaaaagctcagttgggcacctcagaatctgaaccagtc
L A T K A Q L G T S E S E P V
91 ttagatgtctatggcaacaaggctcgaatccaacctcgagtataag
L D V Y G N K V E S N L E Y K
136 cttctcgaagtcaaaaatggcactagtggtggattcagttatacat
L L E V K N G T S G G F S I H
181 ggaggcacaagtgagagtggtccggttgacggttgccaactaagt
G G T S G E C P L D V V Q L S
226 tctcctacggaaagggatcattatgtaagattattaccgtttgac
S P T E R D H Y V R L L P F D
271 aattctacagttgtccgggaatccactgatctcaacttgatagaa
N S T V V R E S T D L N L I E
316 ccagtatggacagttggaaactataatgattcattaggaaaatgg
P V W T V G N Y N D S L G K W
361 tttctaacaaccggcggaatcatcggccaccaggagctaaaacg
F L T T G G I I G H P G A K T
406 ttgctaaattggtttaagcttgagaaagttagtggtcagtataatc
L L N W F K L E K V S V S I Y
451 aatcttgttcattgtccatctggttgcgattcttgcggttaaacctc
N L V H C P S V C D S C V K L
496 tgtaacaaagtcggcatcttctatgtggacggcggttcgacgtcta
C N K V G I F Y V D G V R R L
541 gttctgggttcgtgatgatgatcagcctctgcggttggtgttattt
V L V R D D D Q P L R L V L F
586 ccagctcctaataccttcaagatcctctatgtaa 618
P A P N P S R S S M *

```

**Figura 1.** Sequências de nucleotídeos e aminoácidos deduzida de um gene *Miraculina* de *Citrus sinensis* cv. ‘Valência’, selecionado a partir do banco de dados CAP, identificado como unigene Cs1g09420.1, representando uma ORF. Destaque em verde indica o códon de iniciação da tradução e em róseo indica o códon de parada da tradução.

### 5.3 Estudo da expressão de genes *miraculina* de *Citrus sinensis* cultivares ‘Valência’ e ‘Pêra’

O estudo da expressão de genes *Mir* de *C. sinensis* foi feito tendo por base os dados disponíveis nos dois bancos de dados, o CAP e o CitEST. Com base no sequenciamento do transcriptoma de quatro tipos de tecidos de laranja doce dos órgãos calo, flor, folha e fruto, a análise da expressão dos genes *Mir* de *C. sinensis* cv. ‘Valência’, disponível no banco de dados CAP, revelou que 4 genes apresentaram os mais elevados níveis de expressão, entre os 16 genes encontrados (Figura 2).



**Figura 2.** Genes *Mir* de *C. sinensis* cv. ‘Valência’ com mais alto nível de expressão, de acordo com os dados do banco de dados CAP.

Com relação aos níveis de expressão, entre os genes, o gene Cs5g16780.1 apresentou os mais elevados níveis, com valores variando entre 1.28081 RPKM em calo e de 348.000 RPKM em folha. Por outro lado, entre os órgãos, a mais baixa expressão foi observada em calo para os 4 genes.

Com relação ao perfil de expressão, entre os 4 genes, o perfil foi diferente para os órgãos flor, folha e fruto, indicando uma possível especificidade de expressão em tecido específico. Para o gene Cs1g09420.1, sua mais elevada expressão foi em fruto, tendo sido o único com este perfil entre os 4 genes da Figura 2. Além disso, a expressão desse gene em folha foi 10 vezes menor do que a de fruto. Já para os genes Cs5g16780.1 e Cs5g16770.1 o perfil de expressão foi similar nos órgãos folha e flor, embora o gene Cs5g16780.1 tenha apresentado níveis de expressão mais elevados em ambos os órgãos, como já citado antes. E para o gene Cs5g16850.1 o perfil de expressão foi mais elevado em flor, embora com nível mais baixo do que os genes Cs5g16780.1, Cs5g16770.1.

Observou-se que o gene Cs1g09420.2 apresentou níveis e perfil de expressão idênticos aos do gene Cs1g09420.1, com valores em RPKM também idênticos. A análise de similaridade de sequências via BLASTx revelou que ambos apresentaram 100% de

similaridade com a mesma sequência de *Citros clementine* (Hypothetical Protein CICLE\_v10002395mg), só que com tamanhos diferentes de fragmentos similares (Tabela 2). Então, a partir de um alinhamento dois-a-dois com as sequências de nucleotídeos e de aminoácidos destas duas sequências, observou-se que a sequência de nucleotídeos da ORF do gene Cs1g09420.1 está inteiramente dentro da sequência da ORF de Cs1g09420.2, com 100% de identidade entre elas. A diferença entre elas é que a proteína deduzida a partir de Cs1g09420.2 possui 09 aminoácidos a mais na região carboxi-terminal do que a predita pelo gene Cs1g09420.1, como já descrito na Tabela 3. Uma vez que o gráfico da expressão gênica foi idêntico para os dois genes, sugere-se que os dois genes podem ser idênticos e que: 1) estejam no genoma da laranja doce ‘Valência’, como genes parálogos, como resultado de duplicação gênica, portanto localizados em 2 *loci* gênicos; ou que 2) estejam localizados em apenas 1 *locus* gênico, mas que possam ter ocorrido erros durante a clusterização.

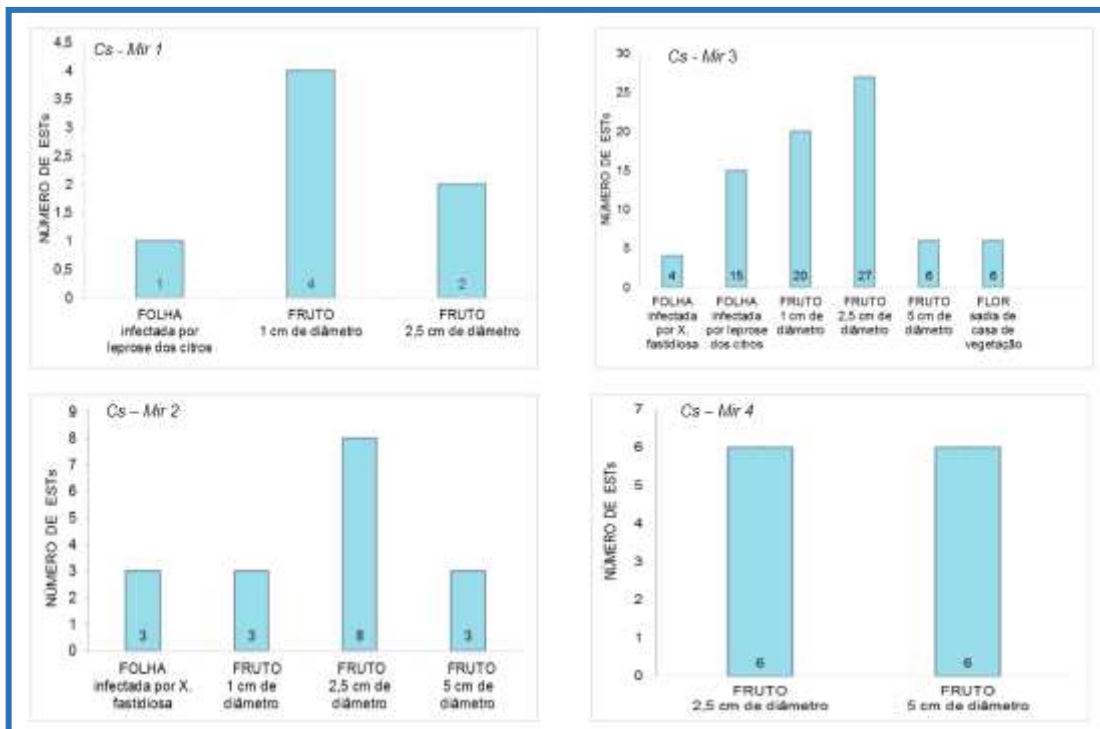
Os outros 11 genes da laranja doce ‘Valência’, cujos gráficos não foram mostrados, apresentaram um baixo nível de expressão, sendo que Cs2g16250.1, Cs5g13890.1, Cs5g13960.1 e Cs5g16920.1 apresentaram níveis variando de 0 a 1,8 RKPM em todos os órgãos. Já os genes Cs5g13930.1, Cs5g16960.1 e as 5 sequências denominadas singlets apresentaram níveis de expressão dentro de uma escala bem mais baixa, variando de 0 a 0,4 RKPM em todos os órgãos, sendo que calo apresentou valores próximos a 0.

A descoberta desses 4 genes *Mir* da laranja doce cv. ‘Valência’, que se mostraram relevantes para estudos posteriores, levou à identificação desses mesmos genes no genoma da cv. ‘Pêra’, por meio de alinhamentos múltiplos e alinhamentos dois-a-dois. Como resultado, os 4 genes foram renomeados como genes únicos de *C. sinensis* (Tabela 4).

**Tabela 4:** Identificação única de genes *Miraculina* de *Citrus sinensis* isolados das cultivares ‘Valência’ e ‘Pêra’.

Identificação única do gene	Cultivar ‘Valência’	Cultivar ‘Pêra’
<i>CsMir1</i>	Cs1g09420.1	CAS-CS-106095
<i>CsMir2</i>	Cs5g16780.1	CAS-CS-107308
<i>CsMir3</i>	Cs5g16770.1	CAS-CS-103575
<i>CsMir4</i>	Cs5g16850.1	CAS-CS-107757

A análise da expressão desses 4 genes *Mir* na cultivar ‘Pêra’ foi feita com base no número de ESTs encontradas nos Contigs dos unigenes indicados na Tabela 3, observando-se em quais órgãos e condições os genes haviam sido expressos (Figura 3).



**Figura 3.** Expressão de 04 genes *Mir* em folha, fruto e flor de laranja doce cv. ‘Pêra’, equivalentes a 04 genes encontrados na cv. ‘Valência’, com base nas informações de ESTs presentes nos contigs disponíveis no banco de dados CitEST.

Com relação aos níveis de expressão, entre os genes de laranja doce ‘Pêra’, o gene *CsMir3* foi encontrado mais expresso em bibliotecas de fruto com diâmetros de 1cm, 2,5 e 5cm, totalizando 52 ESTs encontradas nesse órgão. Além de fruto, o gene *CsMir3* também foi encontrado expresso em folhas infectadas por *Xylella fastidiosa* e pelo vírus da leprose dos citros (4 e 15 ESTs, respectivamente) e em flor (6 ESTs). Por outro lado, entre os órgãos, a frutos (1 a 5 cm de diâmetro) foi observada para os 4 genes, enquanto que o perfil de expressão variou entre eles. Estes dados diferem, em parte, dos observados na cultivar ‘Valência’, provavelmente devido as condições de estresses ambientais na montagem de experimentos para a construção das bibliotecas.

Em relação à expressão dos demais genes *miraculina* de *C. sinensis* cv. ‘Pêra’, e de modo geral, todos foram encontrados predominantemente expressos somente em folhas infectadas por leprose dos citros e *X. fastidiosa*, flor e em casca sadias de casa de vegetação e frutos com diâmetros de 1cm, 2,5cm, 5cm e 7cm, sendo mais expressos frutos na fase inicial de desenvolvimento com 2,5 cm de diâmetros, apesar de no banco de dados CitEST conter sequencias que foram expressas em tecidos de vários órgãos de *C. sinensis* sob diversas condições bióticas e abióticas (REIS et al., 2007; TARGON et al., 2007),

## 5.4 Caracterização molecular de proteínas miraculinas de *Citrus sinensis*

Estudos proteômicos são cruciais para a compreensão da função celular. Neste sentido, a utilização de preditores pode fornecer a localização de regiões importantes para as funções das proteínas de diferentes organismos procaríotos ou eucariotos. Os dados de proteômicas preditos para locais funcionais de anotação de família e eventos pós-traducionais foram indicados de maneira comparativa para as quatro proteínas preditas dos quatro genes encontrados idênticos em ‘Valência’ e ‘Pêra’ na Figura 4.



**Figura 4:** Alinhamento múltiplo das quatro sequências de aminoácidos deduzidas de *C. sinensis* das cultivares “Valência” e “Pêra”. Em negrito a sequência N-terminal de peptídeo sinal, sombreado em amarelo indica a anotação para domínio inibidor de tripsina tipo Kunitz, sombreado em azul os pontos de fosforilação. Sublinhado indica sítios de N-glicosilação, sombreado em cinza indica as cisteínas conservadas, asteriscos e setas pretas indicam as cisteínas que formam pontes dissulfeto.

A predição de sítios de domínios funcionais de proteínas usando do Programa ProSite revelou que somente três proteínas miraculinas de *C. sinensis* (CsMir1, CsMir2 e CsMir4) apresentaram possíveis domínios para Super família de Inibidor de proteinase Tripsina do tipo Kunitz de soja, conforme destacado em amarelo na Figura 4. A maioria das proteínas

miraculinas possuem ação inibidora de tripsina, tendo sido contatada para a Miraculina nativa de *Murraya Koenigi* por GAHLOTH et al. (2011 e 2009). Os inibidores de tripsina da planta de soja do tipo Kunitz (STI) são proteínas com uma massa molecular de aproximadamente 20 kDa e tem geralmente quatro resíduos de cisteínas dispostos em duas pontes dissulfeto constituída em sua fase madura por 190 aminoácidos. Segundo Selvakuman et al. (2011), além desses inibidores desempenharem ação na defesa de plantas contra predadores, são conhecidas por estarem envolvidos em muitas funções biológicas, tais como a coagulação do sangue, agregação de plaquetas e anticarcinogênese.

Como eventos pós-traducionais, a presença de um sítio de clivagem de protease indicando um peptídeo sinal foi confirmado para as 04 proteínas analisadas, o qual está destacado em negrito nas respectivas sequencias alinhadas, apresentando comprimentos de 17, 22, 25 e 29 nt para as proteínas CsMir1 a CsMir4, respectivamente (Figura 4). A sequência do peptídeo sinal é um sítio que indica o direcionamento da proteína para outra organela ou meio extracelular, por meio de proteases o peptídeo sinal é liberado e o restante da sequência representa a proteína madura (BENDTSEN et al., 2004).

Devido a importância do dobramento correto das proteínas sintetizadas para as suas funções, a presença de resíduos de cisteínas conservados dentro da família de proteínas miraculinas foi localizado nas quatro proteínas Mir de *C. sinensis* selecionadas. Como resultado, somente as proteínas CsMir3 e CsMir4 apresentaram 6 resíduos conservados implicados em 03 ligações de pontes dissulfeto (indicadas com asterisco e setas, Figura 3), descritas para miraculina nativa por GAHLOTH et al. (2009). A proteína CsMir1 apresentou 5 resíduos conservados de cys, porém apresentou as quatro cisteínas conservadas descritas para os inibidores de proteinase Tripsina (STI) envolvidas em duas pontes dissulfeto. Já a proteína CsMir2 apresentou 7 cisteínas, sendo apenas 5 em posições esperadas para a formação de pontes dissulfeto, e duas em posições extras, as quais podem interagir com outras moléculas.

Pesquisas tem revelado a importância desse aminoácido (cisteína) para atividades metabólicas dos seres vivos, este desempenha importante papel na estabilidade molecular e resistência a proteólise ao formarem ligações dissulfeto (FANTINI et al., 2008; SANTOS et al., 2010). Resíduos de cisteína e pontes dissulfeto auxiliam na dobragem e estabilidade estrutural de proteínas. A substituição desses resíduos formam ligações dissulfeto com outros aminoácidos e leva a desestabilização proteica (KEISUKE ITO et al., 2011).

A presença de possíveis sítios de N-glicosilação foi confirmada para as quatro proteínas preditas, com número variável, os quais estão indicados sublinhados na Figura 4. A

N-glicosilação ocorre em algumas proteínas através da adição de oligossacarídeos na cadeia proteica em posições específicas e somente no retículo endoplasmático e no complexo de golgi, esse processo ocorre através da ligação N-glicosilação no grupo NH<sup>2</sup> da asparagina, ocorre apenas em motivos Asn-xaa-Ser/Thr onde xaa é qualquer aminoácido exceto prolina (GRUPTA, 2004; KERBAUY, 2008).

Dessa forma, CsMir1 apresentou 4 possíveis sítios de glicosilação nas posições 51 (Asn, Gly, Thr, Ser), 91 (Asn, Ser, Thr, Val), 114 (Asn, Asp, Ser, Leu) e 199 (Asn,-Pro, Ser, Arg). CsMir2 apresentou 3 sítios nas posições 93 (Asn, Ser, Thr, Ile), 128 (Asn, Ala, Ser, Leu) e 180 do tipo (Asn, Gly, Ser, Cys) respectivamente. Segundo Kerbauy (2008) existem casos em que a N-glicosilação ocorre dentro de Asn-Xaa-Cys (uma cisteína que se opõem a uma serina / treonina na posição N + 2). CsMir3 apresentou apenas um sítio de glicosilação na posição 155 do tipo (Asn, Phe, Ser, Tyr), enquanto que CsMir4 apresentou 2 possíveis sítios de glicosilação nas posições 94 e 95 da sequência de nucleotídeo ( Asn, Asn, Ser, Thr) e (Asn, Ser, Thr, Ile).

As células realizam uma série de processos para seu bom funcionamento desde a sinalização, o reconhecimento de moléculas que proporcionem interações ao processo de produção de energia. A fosforilação e a miristoilação podem ter variados propósitos como reconhecimento e adesão celular, sintetização de ATP. Por esta razão, possíveis sítios de fosforilação e de miristoilação também foram estudados.

De acordo com o programa NetPhos, sítios de fosforilação serinas, treonina e tirosina em proteínas eucarióticas (BLOM, 1999) foram evidenciados para as quatro proteínas Mir de *C. sinensis* selecionadas e apontados em azul na Figura 4. Na proteína CsMir1 foram encontrados 13 locais de fosforilação distribuídos ao longo da sequência de aminoácidos, onde 09 são do tipo Ser, 02 do tipo Thr e 2 do tipo Tyr. A presença de 13 sítios de fosforilação foi indicada para a proteína CsMir2, sendo 11 do tipo Ser e 2 do tipo Thr. Para CsMir3 foram encontrados 11 possíveis sítios de fosforilação, sendo 7 do tipo Ser, 1 do tipo Thr e 3 do tipo Tyr (4, 125, 158). Similarmente, a proteína CsMir4 tem em sua constituição 12 sítios de fosforilação, sendo 7 do tipo Ser, 4 do tipo Thr e 1 do tipo Tyr. A fosforilação é o processo metabólico de síntese de ATP a partir da energia liberada pelo transporte de elétrons na cadeia respiratória (SARBASSOV et al., 2005).

Já a predição de sítios de miristoilação N-terminal de proteínas, onde apenas glicinas são miristoiladas (BOLONGA ET AL., 2004), revelou que somente três proteínas miraculinas de *C. sinensis* (CsMir1, CsMir2 e CsMir3) apresentaram possíveis sítios de miristoilação, 01 sítio altamente confiável e 02 de confiabilidade média, em cada uma delas (Dados não



mostrados). A predição para a CsMir4 resultou em dados de baixa confiabilidade com resultados negativos para miristoilação. Através do processo de N-miristoilação proteica ocorre a transferência do ácido mirístico, um ácido graxo saturado de 14 carbonos. De acordo com Gahloth et al. (2011) e Santos et al. (2013), esta ação está relacionada com a viabilidade celular, visto que pode permitir a associação de proteínas com membranas celulares ou interação proteína-proteína.

Os dados de parâmetros físico-químicos analisados para proteínas CsMir maduras, ou seja sem o peptídeo sinal, estão apresentados na Tabela 5.

**Tabela 5:** Identificação de parâmetros físico-químicos de proteínas CsMir maduras

Proteína madura	Comprimento (aminoácidos)	Massa molecular (Da)	pH teórico	Meia vida/tempo		
				Reticulócito humana <i>in vitro</i>	Levedura <i>in vivo</i>	<i>E. coli in vivo</i>
Mir1	185	20398.2	5.55	0,8 h	10 min	10 h
Mir2	215	23412.7	7.46	7,2	>20 h	>10h
Mir 3	213	23194.4	5.84	20h	30min	>10h
Mir4	195	21273.1	5.37	100h	>20h	>10

As quatro proteínas Mir maduras de *C. cinensis* Mir1, Mir3 e Mir 4 apresentaram comprimento de 185, 213 e 195 aminoácidos, com massa molecular de 20398.2, 23194.4 e 21.273.1 Da, respectivamente, enquanto que Mir2 tem comprimento de 215 aminoácidos, com massa molecular de 23412.7 Da. A massa molecular das proteínas madura estar de acordo com o que muitos autores descrevem (PALADINO et al., 2007; SELVAKUMAN, 2014).

O pH teórico das proteínas CsMir1, CsMir3 e CsMir4 se mostrou ácido, variando entre 5.37 a 5.84, enquanto que o da proteína CsMir2 foi de 7,46, a única a apresentar pH teórico básico. A variável de pH pode indicar o local de trabalho da proteína, podendo ser intra ou extra celular. Isomorfas ácidas trabalham fora da célula, podendo estar livres ou acopladas. Pesquisas confirmam que em pH ácido há uma maior expressão do genes *Mir* e que a ativação da propriedade de modificar o sabor azedo em doce depende do meio ao qual encontra-se (MISAKA, 2013).

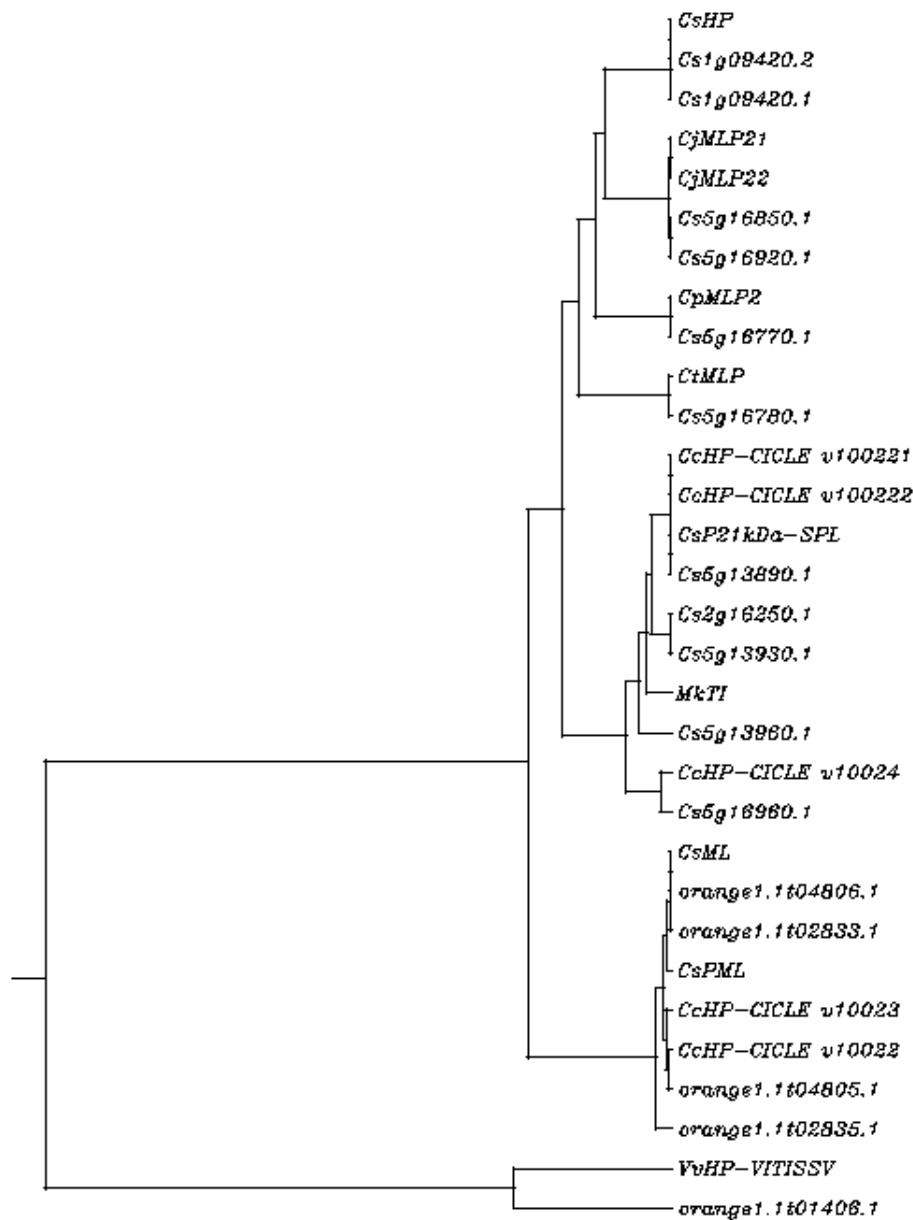
O tempo de meia vida apresentado por elas estima-se entre 0,8 a 100 horas em reticulócito humano *in vivo*, de 10 min a mais que 20 horas em levedura *in vivo* ou seja menor tempo de meia vida, em *Escherichia coli* com 10 ou mais que 10 horas. A meia vida é uma informação relevante para a produção de proteínas heterólogas, visando sua produção

comercial em larga escala. Em alguns organismos como levedura o tempo de vida pode influenciar no perfil de expressão de genes *Mir* (TANASE, 2010).

### **5.5 Inferências filogenéticas de miraculinas de *Citrus sinensis* cultivar ‘Valência’**

Inferências filogenéticas foram dadas a partir dos dados de um alinhamento múltiplo das 16 proteínas CsMir da cultivar ‘Valência’ com 13 sequências similares, sem redundância, utilizando o programa Clustalw, que gerou uma árvore filogenética (dendrograma de similaridade) enraizada com comprimento de braço pelo método de UPGMA (de *Unweighted Pair Group Method with Arithmetic Mean*). A análise filogenética agrupou as sequências em 2 grandes ramos e indicou que todas possuem um mesmo ancestral comum e teriam divergido ao longo do tempo (Figura 5).

Um ramo da árvore foi composto apenas pelo gene Orange1.1t01406.1 e sua sequência similar a VvHP-VITISSY, de *Vitis vinifera*, e representa um ramo diferente das demais espécies agrupadas partindo de um mesmo ramo da árvore filogenética, provavelmente por possuir particularidades não comunicantes entre elas, especialmente devido ao comprimento, como já demonstrado nas Tabelas 2 e 3 (Figura 5).



**Figura 5.** Dendrograma de proteínas miraculina de *C. sinensis* cv. 'Valência' com sequências similares identificadas por BLASTx no GenBank (NCBI), identificadas na Tabela 2.

O segundo ramo da árvore é composto por três subgrupos sendo um constituído por *C. sinensis*, *C. japônica*, *C. tamurana* e Miraculinas-like, levando a crer que compartilham de características em comum. O segundo subgrupo é formado em sua maioria por *C. sinensis* e *Murraya koenigii*, o que corrobora com resultados obtidos em estudos realizados por Gahloth (2009) e Selvakumar (2011), onde afirmaram que *Murraya koenigii* e citros divergem de uma mesma ancestralidade e que estariam próximas evolutivamente.

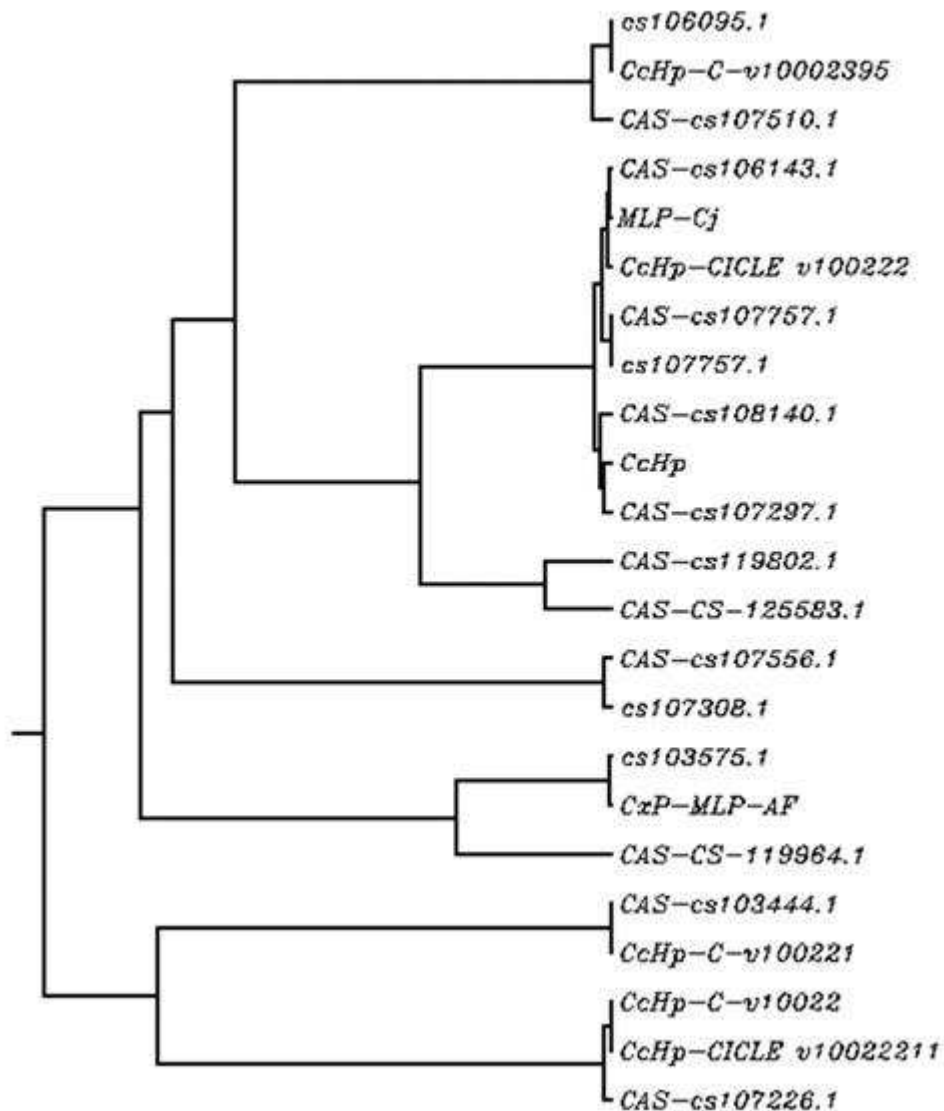
O terceiro subgrupo é formado por proteínas hipotéticas *C. sinensis* e Miraculina-like, podem apresentar caracteres particulares que as classifica em subgrupo diferente.

Estudos filogenéticos indicam que o gênero *Citrus* é bastante antigo, teria divergido por volta de 85 milhões de anos atrás a partir de *Malvales* estando perto filogeneticamente de mamão, cacau e *Arabidopsis* (XU et al., 2014). Um ciclo reprodutivo pouco frequente em alguns taxa por causa da apomixia, do sexo masculino e/ou esterilidade feminina, juvenilidade longa e propagação vegetativa podem contribuir para a restrição da expansão do genoma e evolução. Selvakumar (2011) e Xu et al. (2014) apontaram que o tamanho e a frequência das famílias de genes parálogos acumuladas em cada genoma servem como registro de processo evolutivo.

Fatores genéticos como duplicação gênica, mutações naturais entre outros podem estar influenciando diretamente na expressão de genes *Mir* em *C. sinensis*. A complexidade filogenética e taxonômica dos citros devido a particularidades reprodutivas, caracteres indicadores de parentesco, fundamentados em sua morfologia, anatomia e bioquímica bem como origens de cultivos, confere dificuldade na classificação ao nível de família e subfamília desse gênero (NUNES et al., 2005). A intercompatibilidade sexual possibilita a origem natural de híbridos intergenéticos ao longo do processo evolutivo do grupo (PEREIRA, 2012). Assim a similaridade existente entre miraculinas de *C. sinensis* com demais espécies aqui identificadas podem estar relacionada ao compartilhamento de um ancestral comum com a superfamília Kunitz já que a *Mir* é pertencente a esta superfamília.

### 5.6 Inferências filogenéticas de miraculinas de *Citrus sinensis* cultivar ‘Pera’

A partir de informações atualizadas das sequências *Mir* cv. ‘Pera’, foi feito alinhamento múltiplo utilizando o programa Clustalw (Dados não mostrados), com base nesses dados inferências filogenéticas foram atribuídas para 15 sequências que mostram ORF completa consideradas na análise via BLASTx e 8 similares sem redundância, gerando uma árvore filogenética (dendrograma de similaridade) enraizada com comprimento de braço pelo método de UPGMA (de *Unweighted Pair Group Method with Arithmetic Mean*). A análise filogenética agrupou as sequências em dois grupos próximos de seu ancestral comum (Figura 6).



**Figura 6.** Dendrograma de proteínas miraculina de *C. sinensis* cv. ‘Pera’ com sequências similares identificadas por BLASTx no GenBank (NCBI).

O primeiro grande ramo é formado por 13 unigenes e 5 similares, que possivelmente estariam próximas evolutivamente e teriam divergido a muito tempo considerando a distância entre o ancestral comum e a posição atual evidenciada pelo dendrograma mostrando o compartilhamento de caracteres intraespecíficos, assim possíveis ortólogos.

O segundo ramo é constituído por 2 unigenes e 3 proteínas hipotéticas de *Citrus Clementina*. Estudos moleculares revelam a complexa classificação filogenética dos *Citrus* devido hibridações naturais interespecíficas, mutações espontâneas, alterações cromossômicas estruturais, apomixia e cultivo em ampla área geográfica (RADMANN, 2005).

Alguns autores consideram os genes *Mir* como constituintes de famílias multigênicas, semelhantes quanto a estrutura e funcionalidade. (NEI e RONEY, 2005). Miraculinas caracterizadas mostram papel importante na defesa de plantas através de sua atividade inibidora de tripsina (TSUKUDA et al., 2006). A literatura descreve que ao longo dos tempos inibidores de tripsina evoluíram de um ancestral comum por duplicação de genes, e por eventos de conversão adaptaram-se a esse papel (SELVAKUMAR et al., 2011). Evidências evolutivas indicaram que a maioria das espécies de plantas passou por duplicação de todo o genoma, famílias de genes parálogos serve de registro desse processo evolutivo (WANG et al., 2014).

A miraculina além de possuir atividade inibidora contra ataques de patógenos possui a propriedade modificadora de sabor azedo em doce por ligar-se a receptores específicos de sabor doce (hT1R2-hT1R3) presentes nas papilas linguais humana (MISAKA, 2013). Estudos vem sendo realizado com a utilização desta proteína na alimentação de pacientes com câncer mostrando ter resultados positivos em relação a melhoria nutricional dos pacientes (WILKEN et al., 2012).

## 6 CONCLUSÃO

A partir dos resultados obtidos neste trabalho, é possível concluir que:

- 1) No genoma da espécie vegetal *Citrus sinensis* existem 16 genes *Mir* identificados na cv. “Valência” e disponíveis no banco de dados CAP (*Citrus sinensis Annotation Project*);
- 2) Dez genes *miraculina* da espécie *Citrus sinensis* do genótipo cultivar ‘Valência’ possuem sequências diferentes daquelas do genótipo cultivar ‘Pêra’ selecionadas no banco de dados CitEST;
- 3) Dos 16 genes *Mir* de laranja doce ‘Valência’ identificados, quatro são idênticos a quatro genes isolados da cultivar ‘Pêra’ (ortólogos) e apresentam alto nível de expressão em frutos, folhas e flor sob variadas condições;
- 4) Das quatro proteínas deduzidas de *C. sinensis* selecionadas e caracterizadas molecularmente, três apresentam domínio para Inibidor de Tripsina da superfamília Kunitz;
- 5) Das quatro proteínas deduzidas de *C. sinensis* selecionadas, três apresentam pH teórico ácido e uma básico;
- 6) Genes *CsMir1* e *CsMir2* representam bons candidatos para ensaios de atividade molecular *in silico*, por apresentarem melhores informações moleculares dentre os analisados viáveis ao seu funcionamento a nível de expressão.
- 7) Perspectivas de que novos trabalhos sejam realizados a partir desses resultados.

## 7 REFERÊNCIAS

ALVES, S.M. A bioinformática e sua importância para a biologia molecular. **Revista Brasileira de Ciência e Saúde**. Pombal, PB, v. 3, n. 4, p. 18-25, out./dez., 2013.

ALVINO, Francisco Cassio Gomes et al. Implantação de pomar de citros irrigado no município de Pombal no Estado da Paraíba com duas variedades copas. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, n.1, p. 01–11. Pombal, PB. jan./dez., 2011.

ALTSCHUL, S.F et al. "Basic local alignment search tool." **Journal of Molecular Biology**. p. 403-410, 1990.

ARAÚJO, Eugênio et al. O biólogo das telas de computador. **Boletim do Conselho de Informações Sobre Biotecnologia**, ano 2, n. 7, nov. 2004.

ARAUJO, P. Análise do padrão de expressão dos genes da família MADS-box durante o desenvolvimento do fruto de *Citrus sinensis*. 2010. 51 p. Tese (Mestrado em Biologia Vegetal) Universidade Estadual de Campinas, Campina,SP, 2010.

AZEVEDO, C. L. L. Sistema de Produção de Citros para o Nordeste. **Embrapa Mandioca e Fruticultura Sistema de Produção**, Versão eletrônica Dez/2007.

BACHCHU, Mohammad Adnan Al et al., Functional Expression of Miraculin, a Taste-modifying Protein, in Transgenic Miyagawa Wase Satsuma Mandarin (*Citrus unshiu* Marc.). **Journal of the Korean Society for Applied Biological Chemistry**, v. 54, n.1, p. 24-29, February 2011.

BARRETT, H.C., RHODES, A.M. A numerical taxonomic study of affinity relationships in cultivated *Citrus* and its close relatives. **Systematic Botany**, Notre Dame, v.1, n.2, p. 105-136, 1976.

BASTIANEL, M et al. Caracterização De Genótipos De *Citrus* Spp.Através De Marcadores RAPD. **Ciência Rural**, Santa Maria, v.31, n.5, p.763-768, set./oct., 2001.



BENDTSEN J. D et al. Improved prediction of signal peptides: Signal P 3.0 **Journal of Molecular Biology**, n. 340, p. 783-795, 2004.

BLOM, N., GAMMELTOFT, S., BRUNAK, S. Sequence and structure based prediction of eukaryotic protein phosphorylation sites. **Journal of Molecular Biology**. v. 5, n. 294, 1999.

BOLOGNA, G et al. N-terminal Myristoylation Predictions by Ensembles of Neural Networks. **Proteomics**, v.4. n. 6, p.1626-1632. jun. 2004.

BORGES, H. B. **Redução de Dimensionalidade de Atributos em Bases de Dados de Expressão Gênica**. Universidade Federal do Paraná (Mestrado em Informática). Curitiba, 123p. 2006.

BROUWER, J. N. et al., Miraculin, the Sweetness-inducing Protein from Miracle Fruit **Nature** v.220, p.373-374.1968.

CABALLERO, J.C et al. phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. **Molecular Biology Evolution**, v. 32, n. 4, p. 2015-2035, abril, 2015.

CAICEDO, J. C. C. **Identificação de bactérias do filoplano de citros capazes de interromper o quórum sensing em *Xanthomonas citri* subsp. *citri***. 2015. 68 p. Tese (doutorado) - Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, 2015.

CASAGRANDE, E.C. et al. Expressão gênica diferencial durante déficit hídrico em soja. **EMBRAPA SOJA**, Londrina,PR, agosto de 2001.

CAVALCANTI, L. S. **Papel da celulase XF-0818 na interação *Xylella fastidiosa* X Citros**. Tese de Doutorado, Unidade da USP, Escola Superior de Agricultura Luiz de Queiroz Piracicaba, 2005

CAVALCANTI, L.S. Aspectos bioquímicos e moleculares da resistência induzida. In: CAVALCANTI, L.S. et al. (Eds.). Indução de resistência em plantas Comprehensive Database for Sweet Orange Genome. **PLoS ONE**, v. 9, 2014.

DEL BEM, L. E. V. **Evolução de famílias multigênicas e redes de regulação em plantas**. Tese (doutorado) UNICAMP: Programa de Pós-Graduação em Genética e Biologia Molecular Campinas, SP, 2013.

ENGLER, A. Rutaceae. In: ENGLER, A. & PRANTL, K. (Ed.). **Die Natürlichen Pflanzenfamilien**. 2 ed. Leipzig: Engelmann. v.19, p. 187-359, 1931.

ESPINDOLA, F.S et al. **Recursos de Bioinformática Aplicados às Ciências Ômicas como Genômica**. Instituto de Genética e Bioquímica, Universidade Federal de Uberlândia. p. 27, 2010.

FANTINI, A. P et al. Disponibilidade de ferro em misturas de alimentos com adição de alimentos com alto teor de vitamina C e de cisteína. **Food Science and Technology**, v. 28, n. 2, p. 435-439. Campinas, abr./jun. 2008.

FARIA, L. M. **Mapeamento genético e detecção e QTLs em um cruzamento de limão ‘Cravo’ e citrumelo ‘Swingle**. Instituto Agrônomo de Pós- Graduação. Campinas, SP, p. 01-93, 2007.

FARIAS, A. Q. P. CHACON, P. F. S. DA SILVA, N. R. R. A Bioinformática Como Ferramenta De Formação De Recursos Humanos No IFRN. **Holos**, ano. 28, v. 6, 2012.

FAUS, I. Recent developments in the characterization and biotechnological production of sweet-tasting proteins. **Applied Microbiology and Biotechnology**. v. 53, n. 2, p. 145-151, february 2000.

FÁVERO, P. **Transformação genética de três cultivares de laranja doce a partir de exalantes de plantas adultas**.2010 51 f. Dissertação (Mestrado em Ciências. Área de concentração: Fitotecnia) Piracicaba, p. 01-51, 2010.

GAHLOTH, D et al. Cloning, sequence analysis and crystal structure determination of a miraculin-like protein from *Murraya koenigii*. **Archives of Biochemistry and Biophysics**. v. 494, p. 15–22, 2009.

GAHLOTH, D et al. Bioinsecticidal Activity Of *Murraya Koenigii* Miraculin-Like **Archives Of Insect Biochemistry And Physiology**, v. 78, n. 3, p. 132–144, 2011.

GNANAVEL, M., MUTHUKUMAR, S.P. Identification of novel sweet protein for nutritional Applications. **Bioinformation**, v.7, supl.3, p.112-114, 2011.

GASTEIGER E et al. Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook. **Copyright Humana Press**, p. 571-607, 2005.

BOLOGNA, G et al. N-terminal Myristoylation Predictions by Ensembles of Neural Networks. **Proteomics**, v. 6, n. 4, jun, 2004.

GUINDALINI C., TUFIK S., Uso de microarrays na busca de perfis de expressão gênica - aplicação no estudo de fenótipos complexos **Revista brasileira de psiquiatria** v.29, n.4, São Paulo, 2007.

GUPTA, R.; E. JUNG.; BRUNAK, S. **Prediction of N-glycosylation sites in human proteins In preparation**, 2004. Disponível em (<http://www.cbs.dtu.dk/services/NetNGlyc/>)

HIRAI, T et al. Ubiquitin promoter–terminator cassette promotes genetically stable expression of the taste-modifying protein miraculin in transgenic lettuce. **Plant Cell Reports**, v. 30. p. 2255–2265, 2011.

HIWASA-TANASE, K et al. From miracle fruit to transgenic tomato: mass production of the taste-modifying protein miraculin in transgenic plants. **Plant Cell Reports**, v. 31. p. 513–52, 2012.

IGETA, H. Y et al. Determination of disulfide array and subunit structure of tastemodifying protein, miraculin. **Biochimica et Biophysica Acta**, v. 3, n. 1097, p.303–307, 1991.

INGLETT, G.E., MAY, J.F. Serendipity berries - source of a new intense Interatômica E Metabolômica. **Biosci. J.** Uberlândia. v. 3, n. 26, p. 463-477, may./june, 2010.

INGRAM, J. e BARTELS D. The molecular basis of dehydration tolerance in plants. Annual Review of Plant Physiology and Plant Molecular Biology. v.47, p.377-403, 1996.

ITO, K et al. Microbial production of sensory-active miraculin. **Biochemical and Biophysical Research Communications**. v. 60, p. 407–411. 2007.

ITO, K et al. Cysteine-to-serine shuffling using a *Saccharomyces Cerevisiae* expression system improves protein secretion: case of a nonglycosylated mutant of miraculin, a taste-modifying protein. **Springer Science+Business Media B.V.** 2011.

KATO, K et al. A trial of production of the plant-derived high-value protein in a plant factory. Photosynthetic photon fluxes affect the accumulation of recombinant miraculin in transgenic tomato fruits. **Plant Signaling & Behavior**, v. 6, n. 8, p.1172-1179; august. 2011.

KELLIS, M. B. BRUCE, W. e LANDER, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*, **Nature**, 428, p. 617 - 624, abril 2004.

KERBAUY, W.D et al. Produtos Finais De Glicosilação Avançada (Age) E A Exacerbação Da Doença Periodontal Em Diabéticos – Revisão De Literatura. **Revista de Periodontia**, v. 18, n. 3, setembro, 2008.

KURIHARA, K.; BEIDLER, L. M. Taste-Modifying protein From Miracle Fruit. **Science**, v. 161, n. 3847, p. 1241 - 1243, 20 september 1968.

KURIHARA, Y. Characteristics of antisweet substances, sweet proteins, and sweetness-inducing proteins. **Critical Reviews in Food Science and Nutrition**, v.32, p.231-252, 1992.

KURIHARA, Y.; NIRASAWA, S. Structures and Activities of Sweetness-inducing Substances (Miraculin, Curculin, Strogin) and the Heat-stable Sweet Protein, Mabinlin. **Foods and foos ingredientes jornal of Japan**, v. 174, 1997.

LIFSCHITZ, S. **Algumas pesquisas em bancos de dados e bioinformática**. In: XXVI Congresso da SBC (Workshop de Biologia Computacional), 14-20 jul. 2006, Campo Grande, MS. Anais. Campo Grande: SBC, 2006.

LIMA, C. F. Avaliação do poncirus trifoliata (L) Raf. Como porta enxerto para laramjeira Lima. 2013, 55 f. Dissertação (Mestrado Ciências e Tecnologias Agropecuárias)- Universidade Estadual do Norte Fluminense Darcy ribeiro, Campos dos Goytacazes, RJ. p. 02-54, fevereiro, 2013.

MACHADO, M.A et al. Genética, melhoramento e biotecnologia de citros. In: MATTOS JUNIOR, D., NEGRI, J.D., PIO, R.M., POMPEU, J. **Citros**. Campinas: Instituto Agrônômico e Fundag. p.926, 2005.

MAGALHÃES, D. M. **Análise proteômica de citrumelo ‘swingle’ (*Citrus paradisi* Macfad. Cv. Duncan x *Poncirus trifoliata* (L.) Raf. com alto acúmulo de prolina em situação de déficit hídrico**. 2009, 120 f. Dissertação (Mestrado em Agronomia, Área de Concentração em Produção Vegetal) - Universidade Federal do Paraná, Curitiba, 2009.

MASUDA, Y et al. Cloning and sequencing of a cDNA encoding a taste-modifying protein, miraculin. **Gene**, v.161, p.175-177, 1995.

MELLO, G. E. C. Banco de Dados e Bioinformática. **Boletim de Saúde & Tecnologia**, Unisinos - Universidade do Vale do Rio dos Sinos, n. 25, ano. 5, jan./mar., 2008.

METZGER I. F. et al., Farmacogenética: princípios, aplicações e perspectivas, **Simpósio: Farmacogenética**. v. 39 n.4, p.515-21,2006.

MEYROWINTZ, E. M. Plants compared to animals: the broadest comparative study of development. **Science**, v. 29, p. 1482-1485, 2002.

MISAKA T, Molecular mechanisms of the action of miraculin, a taste-modifying protein. **Seminars in Cell & Developmental Biology**, v.24, n.3 p.222-225, 2013.

MUNIZ, M.F.S., QUEIROZ, F.M., MENEZES, M. Caracterização de isolados de *Phytophthora* patogênicos a *Citrus sinensis* no Estado de Alagoas. **Fitopatologia Brasileira**, p. 201-204, mar./abr. 2004.

NEI, M. ROONEY, A. P. Concerted and birth-and-death evolution of multigene families. **Annual Review of Genetics**, v. 39, p.121-152, 2005.

NEVES, M. F. **Cadeia da laranja já trouxe mais de R\$ 110 bilhões ao país**. Disponível em: <<http://www.zecadirceu.com.br/post.php?id=2289>> acesso em: 10 abril. 2015.

NUNES, Maria Júlia Corazza, et al. Aurantioideae: uma revisão da taxonomia e filogenia, com as contribuições da sistemática molecular. **Cordeirópolis**, v. 26, n 2, p.359-374, 2005.

PROSDOCIMI, Francisco et al. Bioinformática: Manual do Usuário. **Biotecnologia Ciência e Desenvolvimento**, ano. 5, 29 ed. nov./dez. 2002.

OLIVEIRA, J. M. de. **Caracterização morfológica, molecular e patogênica de espécies de *Phytophthora* associadas aos citros**. 2008. 51 p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Lavras, Lavras, 2008.

OLIVEIRA, R.P. RADMANN, E.B. Genetic similarity of citrus fresh fruit market cultivars. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 27, n. 2, p. 332-334, 2005.

OLIVEIRA, R. P. Biotecnologia em citros. **Embrapa Clima Temperado**, 2006. (Embrapa Clima Temperado. Documentos, 160). Pelotas, RS, 2006.

OLIVEIRA, R. R. **Estudo da expressão de genes MADS-box durante o desenvolvimento floral em coffea arábica**. 2011 (?) f. Dissertação (Mestrado em Biologia Vegetal) - Universidade Estadual de Campinas, SP. 2011.

OLIVEIRA, C. R, **Transformação genética de laranjas 'Pera' e 'Natal' (Citrus sinensis L. Osbeck) com o gene atacina A (attA), dirigido por promotores preferencialmente expressos no floema, para resistência a Candidatus Liberibacter spp.** (Dissertação de Mestrado) - Unidade da USP, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2014.

PALADINO, A et al. Functional hypothesis on miraculin' sweetness by a molecular dynamics approach. **Biochemical and Biophysical Research Communications**, v.396, p. 726–730. 2010.

PALADINO, A et al. Molecular modelling of miraculin: Structural analyses and functional hypotheses. **Biochemical and Biophysical Research Communications**, v. 367 p. 26–32, 2008.

PEREIRA, F. D. **Análise Filogenética entre Citrus spp. e Guignardia spp.** Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias (Mestrado) Jaboticabal. p. 436, 2012.

PIO, R.M et al. Variedades copas. In: MATOS JUNIOR, D.; NEGRI, J.D.;PIO, R.M.; POMPEU JUNIOR, J.(Ed). **Citrus**. Campinas: instituto Agrônômico, Fundag, 2005. cap. 3, p. 39-60. 2005.

PROSDOCIMI, F. Introdução À Bioinformática. Curso Online. **Bioteecnologia, Ciência E Desenvolvimento**. 2007. Disponível em: <[http://www2.bioqmed.ufrj.br/prosdocimi/FProsdocimi07\\_CursoBioinfo.pdf/](http://www2.bioqmed.ufrj.br/prosdocimi/FProsdocimi07_CursoBioinfo.pdf/)> acesso em: 05 de out. 2015.

PROSDOCIMI, F et al. Bioinformática: Manual Do Usuário. **Bioteecnologia Ciência & Desenvolvimento**, n. 29, p. 12-25, 2003.

PURUGGANAN M. D et al. Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. **Genetics**, n. 140, p. 345-356, may, 1995.

QIANG XU et al., The draft genome of sweet orange (*Citrus sinensis*), **Nature genetics** v.45, p.59–66, 2013.

RANULFI, A.C. **Utilização de técnicas espectroscópicas no estudo e caracterização de doenças de citros: HLB (greening) e cancro cítrico**. Universidade de São Paulo (Mestrado em Ciências), p. 136, 2014.

RODRIGUES, A.D., Identificação e caracterização *in silico* da expressão de genes do tipo miraculina em *citrus cinensis* 2013. 50 f. Trabalho de Conclusão de Curso – (graduação em Licenciatura em Ciências Biológicas) – Universidade Federal de Campina Grande. 2013.

REIS, M. S et al. Bioinformatics for the Citrus EST Project (CitEST). **Genetics and Molecular Biology**, São Paulo v.30, n.3, 2007.

ROSSETTO, M. P et al., Avaliação da mancha preta dos citros em diferentes variedades de laranja. **Bragantia**, Campinas. v. 70, n. 1, p.58-63, 2011.

SARBASSOV, D. D et al. Phosphorylation and Regulation of Akt/PKB by the Rictor-mTOR Complex. **Science**, v. 307, n. 5712. p. 1098 - 1101, february, 2005.

SANTOS, A.C et al. Organização Supramolecular Da Ftalocianina De Cobalto(Ii) E Seu Efeito Na Oxidação Do Aminoácido Cisteína **Quím. Nova**. v.33, n.3 p. 539-546, São Paulo, 2010.

SANTOS, L.G. Descritores significativos teóricos do sítio de ligação ao substrato da miristoil-coenzima a: Proteína n-miristoiltransferase de candida albicans para o desenho racional de inibidores. **3º Congresso de Iniciação Científica do Centro Universitário do Distrito Federal – UDF** v. 1, n 1, 2013.

SELVAKUMAR, P et al. Structural Insights Into The Aggregation Behavior Ofmurraya Koenigii Miraculin-Like Protein Below Ph 7.5. **Proteins: Structure, Function and Bioinformatics**, v. 82, n. 5, p. 830-840, May, 2014.

SELVAKUMAR, P et al. Molecular Evolution of Miraculin-Like Proteins in Soybean Kunitz Super-Family. **Department of Biotechnology, Indian Institute of Technology**, J Mol Evol. n.73, p. 369–379, 2011.

SIGNALP-3.0, Disponível em: <http://prosite.expasy.org/cgi-bin/prosite/ScanView.cgi> .  
acesso em: 20 nov. 2015.

SIGRIST, C.J.A et al. New and continuing developments at PROSITE. **Nucleic Acids**, 2012.

SILVEIRA, L.; LIBRELOTTO, G. R. Aplicações de XML na bioinformática. **Disc. Scientia**, Série: Ciências Naturais e Tecnológicas, Santa Maria, v. 8, n. 1, p. 107-118, 2007.

SORIANO, L. **Hibridização somática visando à produção de genitores tetraploides para o melhoramento de cultivares copa de citros**. Universidade de São Paulo (Mestre em Ciências). Escola Superior de Agricultura Luiz de Queiros. Piracicaba, 2010.

SUGAYA, T et al. Transgenic strawberry expressing the taste-modifying protein Miraculin. **Plant Biotechnology**, v. 25, p. 329–333, 2008.

SUN, H.J et al. Functional expression of the taste-modifying protein, miraculin, in transgenic lettuce. **FEBS Letters**, v. 580, p.620–626, january, 2006.

SUN H-J. et al., Genetically stable expression of functional miraculin, a new type of alternative sweetener, in transgenic tomato plants **Plant Biotechnology Journal**, v.5, p. 768–777, 2007

SWINGLE, W.T.; REECE, P.C. The botany of citrus and its wild relatives. In: REUTHER, W.; WEBBER, H.J.; BATCHELOR, L.D. (Ed.). The citrus industry. Riverside: **University of California**, v. 1, p.190-430, 1967.

TANAKA, T. Fundamental discussion of *Citrus* classification. **Studia Citrologica**, Fukuoka-ken, v. 14, p.1-6, 1977.

TARGON, M. L. P. N. et al., CitEST libraries **Genetics and Molecular Biology**, v.30, n.3 p.1019-1023, 2007.

TAKITA, M.A et al. CitEST libraries. **Genetics and Molecular Biology**, v. 30, supl. 3, p. 1019-1023, 2007.



TSUKUDA S, et al., Characterization of cDNAs encoding two distinct miraculin-like proteins and stress-related modulation of the corresponding mRNAs in Citrus jambhiri Lush. **Plant Molecular Biology**, v.60 n.1 p.125-136, 2006.

TATUSOV T e TATUSOV R, ORF finder. Disponível em: <<http://www.ncbi.nlm.nih.gov/projects/gorf/>> acesso em 30 de novembro de 2015

THEERASILP S. e KURIHARAS Y, Complete Purification and Characterization of the Taste-modifying Protein, Miraculin, from Miracle Fruit. **The American Society for Biochemistry and Molecular Biology**, v.263, n.23 p. 11536-11539, 1988.

THOMPSON, C. E. **Divergência funcional da família gênica da álcool desidrogenase em plantas**. 2006. 122 p. Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul. Instituto de Biociências. Porto Alegre, RS, junho, 2006.

THOMPSON, J.D., HIGGINS, D.G. GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, n. 22, 1994.

WANG. J et al. Citrus sinensis Annotation Project (CAP): A Comprehensive Database for Sweet Orange Genome. **PLoS ONE**, v. 9, n. january, 2014.

WILKEN, M.K., SATIROFF, B.A. Pilot Study of “Miracle Fruit” to Improve Food Palatability for Patients Receiving Chemotherapy. **Clinical Journal of Oncology Nursing**, v. 16, n. 5, october, 2012.

YANO, M et al. Tomato is a suitable material for producing recombinant miraculin protein in genetically stable manner. **Plant Science**, v. 178, p. 469–473, 2010.

YE, X.Y., WANG H.X., NG, T.B. Sativin: A Novel Antifungal Miraculin-Like Protein Isolated From Legumes Of The Sugar Snap *Pisum Sativum* Var. *Macrocarpon*. **Life Sciences**, v. 67, n. 7, p. 775–781, 7 july, 2000.

YI LI et al., Plant Expansins Are a Complex Multigene Family with an Ancient Evolutionary Origin. **Plant Physiology**, v. 128, p. 854–864, 2002.