

IDENTIFICAÇÃO E CARACTERIZAÇÃO *IN SILICO* DE GENES *MIRACULINA* EM *Citrus sinensis* NAS CULTIVARES ‘PÊRA’ E ‘VALÊNCIA’

Adeilma Fernandes de Souza (1); Erica Renata Nogueira Sá (1); Marco Aurélio Takita (2);
Rafael Trindade Maia (3); Magnólia de Araújo Campos (4)

Universidade Federal de Campina Grande - UFCG

Centro de Educação e Saúde – CES

adeilmafsouza@gmail.com

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 pacientes com doenças relacionadas ao consumo de açúcar, como os diabéticos, isso porque possui baixas calorias e é capaz de adoçar com quantidades bem menores em relação a outros adoçantes já existentes no mercado GNANAVEL e MUTHUKUMAR, 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). O gênero *Citrus* inclui algumas das mais importantes árvores frutíferas cultivadas comercialmente em todo o mundo (CABALLERO, 2015), comercialmente as cultivares ‘Pêra’ e ‘Valência’ de laranja doce (*Citrus sinensis*) ganham destaque.

Para efetivação desta pesquisa utilizou-se informações da expressão gênica de *Miraculina* em de *Citrus sinensis*, disponíveis em dois bancos de dados: o CitEST (*Citrus Expressed Sequence Tags*), do Centro APTA de Citros Sylvio Moreira, Cordeirópolis SP, que trabalha com o sequenciamento da cultivar laranja - doce ‘Pêra’, e no banco de dados chinês CAP (*Citrus sinensis Annotation Project*), que se encontra disponível online, com os dados do sequenciamento do genoma estrutural específico de Laranja – doce cultivar ‘Valência’.

O estudo da expressão do gene *Miraculina* em *Citrus sinensis* é de grande relevância, considerando a importância financeira atribuída às culturas de *Citrus* e variedades estudadas, esse gene está envolvido na resposta da planta ao estresse biótico e abiótico, bem como o

(83) 3322.3222

contato@conapesc.com.br

www.conapesc.com.br

desenvolvimento dos frutos, assim este trabalho tem como principais metas, identificar sequências de genes *Miraculina* de *C. sinensis* nas 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, e fazer inferências filogenéticas a partir de alinhamento das sequências de proteínas.

2. 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. As sequências gênicas estudadas foram obtidas a partir de dois bancos de dados: 1) 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) 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 disponível 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>. Em seguida, 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 dendrograma de similaridade, 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>.

3. RESULTADOS E DISCUSSÃO

3.1 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. 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.

Para compreender a observação do número de unigenes diferentes encontrado nos dois genótipos da mesma espécie, cultivares ‘Valência’ e ‘Pêra’, várias análises foram feitas usando somente as sequências das regiões codificadoras (CDS), na tentativa de encontrar os pares de genes idênticos nos dois genótipos, incluindo: 1) alinhamentos múltiplos de todas as sequências 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 sequências de nucleotídeos, obtidas nos dois bancos de dados, e de aminoácidos deduzidos; e 3) clusterização de todas as sequências 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 sequências 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 sequência

considerada igual (contigs) ou por apenas uma sequência (singlet). Assim, a baixa qualidade de algumas sequências de DNA leva estas a não se agruparem em *Contigs* de maneira correta, o que pode explicar, em parte, o acontecido com as sequências da cv. ‘Pêra’ do banco de dados CitEST.

3.2 Caracterização molecular de genes *miraculina* de *Citrus sinensis*

As sequências de cada um dos unigenes encontrados em ambos os bancos de dados foram submetidas a análises de BLASTx e as sequências 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.

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 sequências), *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 sequências *Mir* de laranja doce ‘Pêra’ foi atualizada levando a confirmação de que as 25 sequências apresentaram similaridade com proteínas *Mir* de outras espécies de citros, sendo somente uma (01) com *Murraya koenigii*. Entretanto, uma análise minuciosa revelou que nove (09) sequências encontradas por palavra-chave *Miraculin* no banco de dados CitEST apresentaram similares apenas de trechos 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), *Citrus japonica* (3), 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 *C. sinensis* não ser uma espécie pura (ARAÚJO e ROQUE 2005, 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 1).

Tabela 1: 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 1, 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*.

Por meio de alinhamentos múltiplos e alinhamentos dois-a-dois, foram identificados dos 16 genes *Mir* de laranja doce ‘Valência’ quatro genes idênticos na cultivar ‘Pêra’ e renomeados como genes únicos de *C.sinensis*. (Tabela 2).

Tabela 2: 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

3.3 Agrupamentos de miraculinas de *Citrus sinensis* cultivar ‘Valência’ por similaridade

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, as proteínas foram agrupadas por similaridade enraizada com comprimento de braço pelo método de UPGMA (de *Unweighted Pair Group Method with Arithmetic Mean*). O dendrograma gerado agrupou as sequências em 2 grandes ramos e indicou que todas possuem um mesmo ancestral comum e teriam divergido ao longo do tempo (Figura 1).

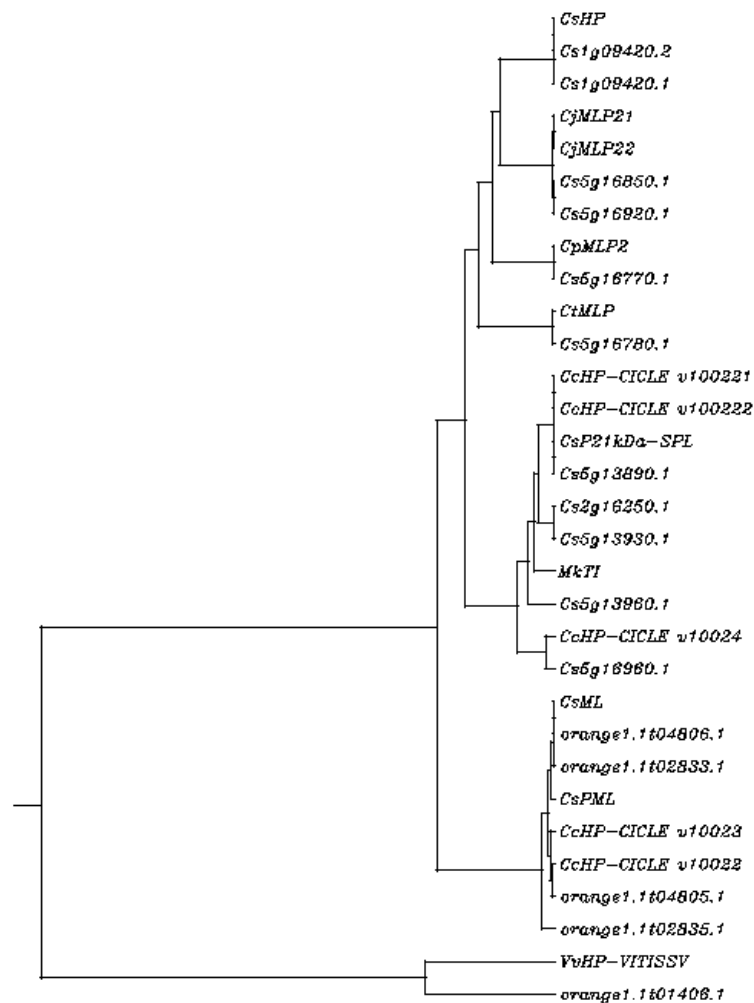


Figura 1. Dendrograma de proteínas miraculina de *C. sinensis* cv. ‘Valência’ com sequências similares identificadas por BLASTx no GenBank (NCBI), identificadas na Tabela 2.

Um ramo do dendrograma 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, provavelmente por possuir particularidades não comunicantes entre elas, especialmente devido ao comprimento, como já demonstrado na Tabela 1 e (Figura 1).

O segundo ramo do dendrograma é composto por três subgrupos, sendo um grupo constituído por *C. sinensis*, *C. japônica*, *C. tamurana* e Miraculin-like, levando a acreditar 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, juvenildade 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 logo 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.

4. CONCLUSÕES

A partir dos resultados obtidos neste trabalho é possível concluir que 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*) e que dez desses genes *miraculina* da espécie *Citrus sinensis* possuem sequencias diferentes daquelas do genótipo cultivar ‘Pêra’ selecionadas no banco de dados CitEST. Dos 16 genes *Mir* de laranja doce ‘Valência’ identificados, quatro são idênticos a quatro genes isolados da cultivar ‘Pêra’, e destes quatro três apresentam domínio para Inibidor de Tripsina da superfamília Kunitz.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO, E.F.; ROQUE, N. Taxonomia dos citros. In: Mattos Júnior, D.; De Negri, J.D.; Pio, R.M.; Pompeu Júnior, J. (eds) Citros. Campinas: Instituto Agronômico e Fundag, p.125-145, 2005.

ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. & LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology** v.215, p.403-410, 1990. Disponível em: <<http://blast.ncbi.nlm.nih.gov/blast/Blast.cgi>>. Acesso em 13 nov. 2015.

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.

Centro APTA de Citros Sylvio Moreira. Disponível em: <<http://biotecnologia.centrodecitricultura.br/>>. Acesso em 15 nov. 2015.

Citrus sinensis Annotation Project. Disponível em: <<http://citrus.hzau.edu.cn/orange/>>. Acesso em: 10 nov. 2015.

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.

HUANG, X. e MADAN, A. (1999) CAP3: A DNA sequence assembly program. **Genome Research**, v.9, p. 868-877. Disponível em: <<http://douw.prabi.fr/software/cap3>>. Acesso em: 6 nov 2015.

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

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.

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.

QIANG XU et al., The draft genome of sweet orange (*Citrus sinensis*), **Nature Genetics**, v.45, p.59–66, 2014.

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

THEERASILP, S.; KURIHARA, Y. Complete purification and characterization of the taste-modifying protein, miraculina, from miracle fruit. **The Journal of Biological Chemistry**. V.263, n.23. p.11536-11539. 1988.

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.

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