

Identificação e caracterização *in silico* de RNAs longos não-codificantes associados a diferentes tecidos e fases de desenvolvimento em tomate

In silico identification and characterization of long non-coding RNAs associated with different tissues and developmental stages in tomato

DOI: 10.34188/bjaerv5n3-041

Recebimento dos originais: 06/05/2022

Aceitação para publicação: 30/06/2022

Leandro de Oliveira Silva

Doutor em Biotecnologia Vegetal pela Universidade Estadual do Norte Fluminense Darcy Ribeiro
Instituição: Universidade Estadual do Norte Fluminense Darcy Ribeiro
Endereço: Av. Alberto Lamego, 2000, Campos dos Goytacazes – RJ, Brasil
E-mail: oliveiradasilvaleo@gmail.com

Juliana Lopes Moraes

Mestra em Biotecnologia Vegetal pela Universidade Estadual do Norte Fluminense Darcy Ribeiro
Instituição: Universidade Estadual do Norte Fluminense Darcy Ribeiro
Endereço: Av. Alberto Lamego, 2000, Campos dos Goytacazes – RJ, Brasil
E-mail: juliabio.diversidade@gmail.com

Deise Ferreira Fernandes Paes

Mestra em Biociências e Biotecnologia pela Universidade Estadual do Norte Fluminense Darcy Ribeiro
Instituição: Universidade Estadual do Norte Fluminense Darcy Ribeiro
Endereço: Av. Alberto Lamego, 2000, Campos dos Goytacazes – RJ, Brasil
E-mail: deisepaes@pq.uenf.br

RESUMEN

Consumido pelos astecas desde 700 d.C., o tomate é nativo das Américas Central e do Sul. Foi domesticado no México e, mais tarde, levado para a Europa, onde era utilizado para decoração. Seja *in natura* ou na forma de molhos e extratos, depois dos tubérculos, o tomate é a hortaliça mais consumida no Brasil. Entender os mecanismos moleculares que atuam no tomateiro em diferentes tecidos e estádios do desenvolvimento pode fornecer as bases para o melhoramento desta planta. O objetivo deste artigo é identificar e caracterizar, por meio de análises de bioinformática, RNAs longos não-codificantes em diferentes tecidos de tomate. A partir de um transcriptoma com 70.432 transcritos, foram identificados 5.668 lncRNAs. Análises de expressão mostraram que diversos lncRNAs têm alta especificidade de tecido e estádio de desenvolvimento. Dez diferentes lncRNAs foram identificados como alvos de 13 miRNAs. Análises de *Gene Ontology* dos alvos do lncRNA MSTRG.24890.1, que possui alta expressão em fruto, indicaram o núcleo como componente celular principal e, como função molecular, a ligação a ácido nucléico. Isso sugere que MSTRG.24890.1 pode interagir com DNA ou RNA. Esperamos que a caracterização de lncRNAs diferencialmente expressos em tomate possa contribuir para uma maior compreensão a respeito das funções destas importantes moléculas reguladoras da expressão gênica.

Palabras clave: *Solanum lycopersicum*, lncRNAs, transcriptoma, bioinformática.

ABSTRACT

Consumed by the Aztecs since 700 AD, the tomato is native to Central and South America. It was domesticated in Mexico and later taken to Europe, where it was used for decoration. Whether *in natura* or in the form of sauces and extracts, after tubers, tomato is the most consumed vegetable in Brazil. To understand the molecular mechanisms that act on tomato in different tissues and stages of development can provide the basis for the improvement of this plant. The objective of this article is to identify and characterize, through bioinformatics analysis, long non-coding RNAs in different tomato tissues. From a transcriptome with 70,432 transcripts, 5,668 lncRNAs were identified. Expression analyzes showed that several lncRNAs have high tissue and developmental stage specificity. Ten different lncRNAs were identified as targets of 13 miRNAs. Gene Ontology analysis of the targets of the lncRNA MSTRG.24890.1, which has high expression in fruit, indicated the nucleus as the main cellular component and, as a molecular function, the binding to nucleic acid. This suggests that MSTRG.24890.1 can interact with DNA or RNA. We hope that the characterization of lncRNAs differentially expressed in tomato may contribute to a better understanding of the functions of these important regulatory molecules of gene expression.

Keywords: *Solanum lycopersicum*, lncRNAs, transcriptome, bioinformatics.

1 INTRODUÇÃO

O tomateiro é originário das Américas Central e do Sul, especificamente das regiões andinas do Peru, Bolívia e Equador. Foi domesticado na região sul do México, onde viviam os astecas (FERREIRA, 2004), que o consumiam desde 700 d.C. Nos Andes ainda é possível encontrar formas primitivas, como *Lycopersicon pimpinellifolium* L. (tomate cereja), considerado o ancestral mais próximo da espécie cultivada atualmente (COELHO et al. 2018).

O tomate atualmente cultivado (*Solanum lycopersicum*) é consumido tanto na forma *in natura* quanto empregado para a confecção de molhos e extratos. É rico em licopeno, uma substância antioxidante. O maior produtor é a China, com mais de 58 milhões de toneladas/ano (em 2016 o Brasil ocupava o 9º lugar, atrás da Índia, Estados Unidos, Turquia, Egito, Itália, Irã e Espanha) (CONAB, 2019).

A tomaticultura é de extrema importância para os municípios brasileiros produtores e para os consumidores, mas é uma atividade que expõe agricultores a riscos e incertezas, pois o tomate sofre muitas flutuações de preço (CARVALHO et al. 2014). De forma similar às demais hortaliças, o tomate perece rapidamente, e há perdas consideráveis desde a colheita até chegar à mesa do consumidor (COSTA; FILHO, 1996). Além disso, estresse abiótico (como seca e salinidade) e biótico (como ataque de pragas) causam redução da produção das lavouras, que são exigentes em termos de fertilizantes e defensivos (PEREIRA-CARVALHO et al. 2014). Entender os mecanismos moleculares que atuam no tomateiro em diferentes tecidos e estádios do desenvolvimento pode fornecer novas bases moleculares para o melhoramento desta planta, buscando a obtenção de cultivares mais produtivos e resistentes.

Uma classe de moléculas reguladoras da expressão gênica que tem se destacado é a dos RNAs não-codificantes (ncRNAs, do inglês *non-coding RNAs*). Em eucariotos, cerca de 90% do genoma é transcrito em ncRNAs. Dentro do grupo dos ncRNAs, os RNAs longos não-codificantes (lncRNAs) têm sido descritos como reguladores de importantes processos biológicos, em plantas e animais. Os lncRNAs têm as seguintes características principais: >200 nt de comprimento e ausência de potencial codificador (WANG; CHEKANOVA, 2017).

Em tomate, lncRNAs já foram identificados e caracterizados em diferentes trabalhos: em plantas infectadas por *Phytophthora infestans*, foi demonstrado que Sl-lncRNA15492 é clivado por Sl-miR482a (JIANG et al. 2020); lncRNAs associados ao amadurecimento do fruto foram identificados utilizando 134 transcriptomas (WANG et al. 2018a); em raiz de tomate submetido a estresse salino (LI et al. 2021); os lncRNAs ncRNA42705 e lncRNA08711, quando silenciados resultam no aumento de miR159 e decréscimo de MYB, e estão associados a resistência a *P. infestans* (CUI et al. 2020); em tomate infectado com vírus da folha amarela de tomate (WANG et al. 2018b), e lncRNAs atuando como *mimic* para miRNA em tomate infectado com vírus da folha amarela de tomate (WANG et al. 2015). Não foram identificados trabalhos caracterizando lncRNAs diferencialmente expressos entre diferentes tecidos e estádios do desenvolvimento do tomate.

A identificação de lncRNAs associados a diferentes tecidos e estádios do desenvolvimento em tomate pode auxiliar em uma maior compreensão dos processos regulados por essas importantes moléculas, contribuindo para a pesquisa e obtenção de novos cultivares. O presente trabalho tem como objetivo identificar e caracterizar, por meio de abordagem computacional, lncRNAs diferencialmente expressos em tomate, utilizando um transcriptoma obtido de diferentes tecidos.

2 MATERIAIS E MÉTODOS

2.1 OBTENÇÃO DAS BIBLIOTECAS DE RNA-SEQ

As análises foram conduzidas por meio do ambiente *UseGalaxy* [<https://usegalaxy.org/>]. Bibliotecas de RNA-seq públicas de tomate (*Solanum lycopersicum* cv Heinz) foram baixadas do *Sequence Read Archive* (SRA) (**Quadro 1**) (Illumina HiSeq 2000; *single end*; *forward*). As bibliotecas foram submetidas ao programa FASTQC (ANDREWS, 2010) para remoção dos *reads* de baixa qualidade (<20). As bibliotecas foram submetidas ao programa Trimmomatic (BOLGER; LOHSE; USADEL, 2014) para remoção dos adaptadores resultantes do processo de sequenciamento.

Quadro 1: Bibliotecas de *RNA-seq* de tomate.

Órgão	Código
Botão de flor	SRR404315
	SRR404316
Flor	SRR404313
	SRR404314
Raiz	SRR404311
	SRR404312
Folha	SRR404309
	SRR404310
Fruto 10 dias após o estágio <i>breaker</i>	SRR404328
	SRR404329
Fruto <i>breaker</i>	SRR404326
	SRR404327
Fruto <i>mature green</i>	SRR404324
	SRR404325
Fruto de 3 cm	SRR404321
	SRR404322
Fruto de 2 cm	SRR404319
	SRR404320
Fruto de 1 cm	SRR404317
	SRR404318

Fonte: Elaborado pelos autores deste artigo.

2.2 MONTAGEM DO TRANSCRIPTOMA

Os *clean reads* obtidos após aplicação dos filtros de qualidade foram empregados para a montagem de um transcriptoma. Os *reads* foram alinhados contra o genoma de tomate (*Slycopersicum_691_ITAG4.0*) baixado do Phytozome (GOODSTEIN et al. 2012). Os transcritos foram montados por meio do programa StringTie (PERTEA et al. 2015). A qualidade da montagem foi analisada por meio de GffCompare (PERTEA, 2020).

2.3 CÁLCULO DA EXPRESSÃO DIFERENCIAL DE TRANSCRITOS

A expressão dos transcritos foi obtida por meio do programa Kallisto (BRAY et al. 2016). Para a expressão diferencial, foram consideradas as seguintes comparações: botão de flor/flor; folha/raiz; fruto de 2 cm/ fruto de 1 cm; fruto de 3 cm/ fruto de 1 cm; fruto *mature green*/ fruto de 1 cm; fruto *breaker*/ fruto de 1 cm; fruto de 10 dias/ fruto de 1 cm. Foram considerados diferencialmente expressos, em cada comparação, os transcritos que apresentaram \log_2 *fold change* (FC) >2 (induzidos) ou < -2 (reprimidos). O *heatmap* de expressão foi obtido por meio do programa Morpheus [<https://software.broadinstitute.org/morpheus/>].

2.4 IDENTIFICAÇÃO DE LNCRNAs NOVOS E DOS CONHECIDOS

Para a identificação dos lncRNAs de tomate, primeiramente foram descartados os transcritos com tamanho menor do que 200 nucleotídeos (nt). Em seguida, os transcritos foram submetidos ao

programa Transdecoder v.5.0.2 [<http://transdecoder.github.io>], e aqueles que apresentaram *open reading frames* (ORFs) maiores do que 300 nt (100 aminoácidos, aa) foram descartados. Os transcritos foram, então, alinhados contra o banco Rfam (GRIFFITHS-JONES et al. 2003), e aqueles que tiveram *hit* com alguma sequência foram descartados. Os transcritos foram, então, alinhados a sequências codificantes (do inglês *coding sequence*, CDS) de tomate, e aquelas que apresentaram alinhamento *plus* com 100% de identidade foram descartadas. Em seguida, os transcritos foram submetidos a diferentes ferramentas para cálculo do potencial codificante: CPC (KONG et al. 2007); CPC2 (KANG et al. 2017); RNASamba (CAMARGO et al. 2020); CNIT (GUO et al. 2019) e DeepPlnc (RITU et al. 2021). As sequências identificadas como lncRNAs pelos cinco programas foram considerados lncRNAs. O diagrama de Venn das sequências consenso foi obtido por meio da ferramenta on-line Venn [<http://bioinformatics.psb.ugent.be/webtools/Venn/>].

Foram baixados 4.716 lncRNAs do banco CANTATA (SZCZESŃIAK; ROSIKIEWICZ; MAKALOWSKA, 2016) e 4.242 lncRNAs do banco GreeNC (GALLART et al. 2015). Os lncRNAs identificados no transcriptoma foram alinhados contra os lncRNAs dos dois bancos de dados, e aqueles que apresentaram 100% de cobertura foram considerados conhecidos. Os lncRNAs foram alinhados contra CDS de tomate (*script* próprio) e classificados de acordo com o alinhamento: senso (*plus*) (menos de 100% de identidade em *plus*), antisenso (*minus*) (alinhamento em *minus*) e lincRNAs (não alinharam).

2.5 CÁLCULO DO CONTEÚDO CG, PREDIÇÃO DA ESTRUTURA SECUNDÁRIA E LOCALIZAÇÃO SUBCELULAR

O conteúdo CG dos lncRNAs e dos mRNAs foi obtido por meio da ferramenta on-line *GC Content Calculator* [<https://jamiemcgowan.ie/bioinf/gc.html>]. A estrutura secundária de lncRNAs selecionados (MSTRG.1361.1 e MSTRG.40151.1, os dois únicos preditos para o núcleo; MSTRG.10925.1, MSTRG.29733.2 e MSTRG.29733.1, com alta expressão apenas em botão floral; MSTRG.26001.1, com alta expressão apenas em raiz e MSTRG.22083.5 e MSTRG.22083.2, com alta expressão em fruto “breaker”) foram obtidos por meio das ferramentas *RNAfold web server* (GRUBER et al. 2008), e a localização subcelular por meio de *DeepLncLoc* (ZENG et al. 2022).

2.6 INTERAÇÃO LNCRNA-MIRNA, INTERAÇÃO LNCRNA-MRNA E ANÁLISE DE *GENE ONTOLOGY* (GO) DOS MRNAS ALVO

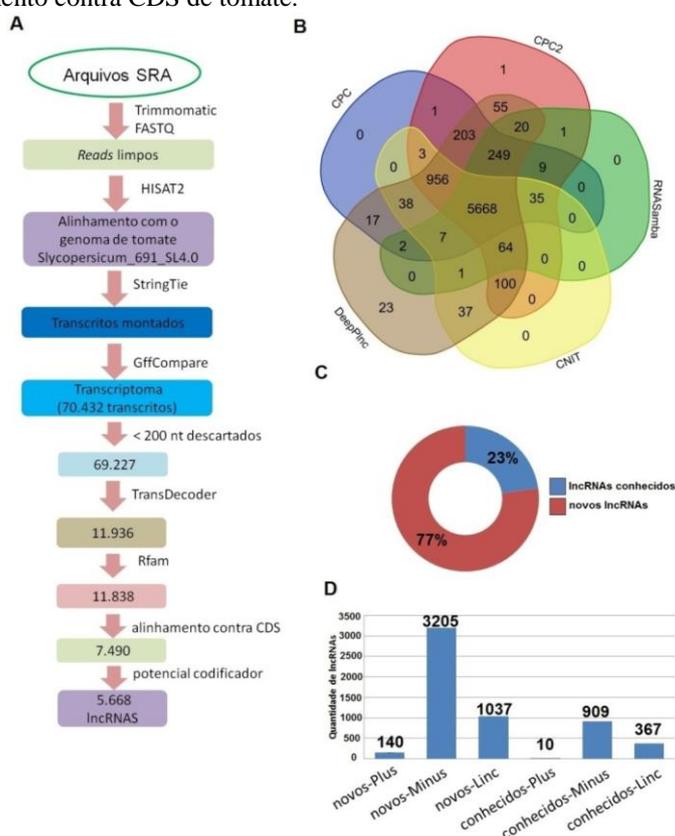
A predição dos lncRNAs que são clivados por miRNAs foi obtida por meio da ferramenta psRNATarget (DAI; ZHAO, 2011). A predição dos RNAs mensageiros (mRNAs) que podem ser alvos de lncRNAs foi feita por meio da ferramenta IntaRNA (MANN; WRIGHT; BACKOFEN,

2017). Foram considerados como alvos dos lncRNAs os transcritos que alinharam com menor valor de energia de hibridização. A análise de GO foi realizada por meio da ferramenta on-line agriGO v2.0 (TIAN et al. 2017). Os 347 alvos com os menores valores de energia de hibridização foram selecionados e submetidos ao programa agriGO v2.0.

3 RESULTADOS E DISCUSSÃO

O *pipeline* utilizado para a identificação de lncRNAs em tomate está representado na **Figura 1-A**. A montagem do transcriptoma permitiu obter 70.432 transcritos, e foram identificados 5.668 lncRNAs. Destes, 1.286 foram considerados lncRNAs conhecidos, e 4.832 novos lncRNAs. Foram considerados lncRNAs as sequências identificadas pelos cinco programas (**Figura 1-B**), totalizando 5.668 lncRNAs. Considerando comparação com os bancos de dados CANTATA e GreeNC, a maioria dos lncRNAs identificados é nova (77%) (**Figura 1-C**). Após alinhamento contra CDS de tomate, percebemos que a maioria dos novos lncRNAs, assim como dos conhecidos, é antisense (orientação *minus*) (**Figura 1-D**). O número de novos lncRNAs intergênicos (1037) é bem superior ao de conhecidos intergênicos (367).

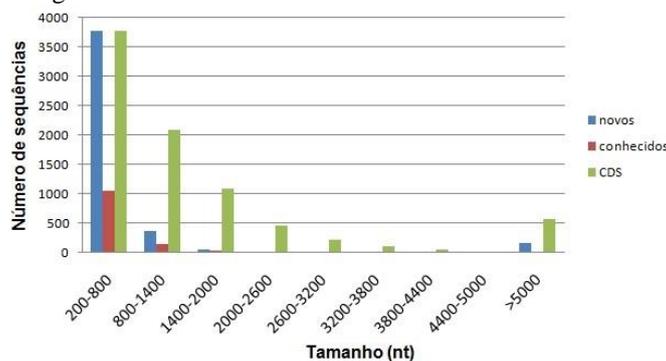
Figura 1: Identificação de lncRNAs em tomate. 1-A – *Pipeline* utilizado para a identificação dos lncRNAs. 1-B – Diagrama de Venn mostrando os lncRNAs identificados pelos 5 programas utilizados para cálculo de potencial codificante. 1-C – Podemos observar que 23% dos lncRNAs são conhecidos e 77% são novos. 1-D – Classificação dos lncRNAs segundo alinhamento contra CDS de tomate.



Fonte: Elaborado pelos autores deste artigo.

Na **Figura 2**, podemos observar a distribuição de tamanho dos lncRNAs em comparação com seqüências codificantes de tomate. Percebemos que a grande maioria dos lncRNAs possui tamanho inferior a 800 nt. Grande parte das CDS possuem entre 800-3800 nt. Uma quantidade considerável de lncRNAs possui tamanho superior a 5.000 nt.

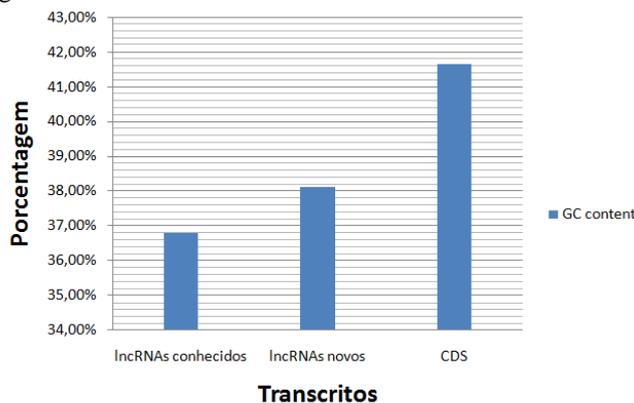
Figura 2: Tamanho dos lncRNAs e dos mRNAs de tomate.



Fonte: Elaborada pelos autores deste artigo.

Na **Figura 3** observamos que os lncRNAs possuem conteúdo GC menor do que as seqüências codificantes. Uma menor proporção de conteúdo GC é perfeitamente condizente com a literatura, e ajuda a explicar sua localização nuclear e baixa expressão. Comparados com transcritos codificantes, lncRNAs possuem menor conteúdo GC (NIAZI; VALADKHAN, 2012).

Figura 3: Conteúdo GC dos lncRNAs identificados em tomate.



Fonte: Elaborado pelos autores deste artigo.

A predição de alvos por meio da ferramenta psRNATarget permitiu identificar 10 diferentes lncRNAs como alvos de 13 miRNAs. Esses miRNAs pertencentes a 6 diferentes famílias (*e-value* máximo de 1,0) (**Quadro 2**). Dos lncRNAs identificados como alvos de miRNAs, quatro são conhecidos e seis são novos.

Quadro 2: LncRNAs identificados como alvos de miRNAs.

LncRNAs	Classificação	MiRNAs	E-value
MSTRG.14093.1	conhecido	sly-miR156a	0,0
MSTRG.14093.1	conhecido	sly-miR156b	0,0
MSTRG.14093.1	conhecido	sly-miR156c	0,0
MSTRG.22083.2	conhecido	sly-miR9470-5p	1,0
MSTRG.22083.4	conhecido	sly-miR9470-5p	1,0
MSTRG.22083.5	conhecido	sly-miR9470-5p	1,0
MSTRG.29733.2	novo	sly-miR156e-3p	0,0
MSTRG.29733.1	novo	sly-miR156e-3p	0,0
MSTRG.6806.1	novo	sly-miR167b-5p	0,0
MSTRG.24581.1	novo	sly-miR172a	0,0
MSTRG.24581.1	novo	sly-miR172b	0,0
MSTRG.37256.1	novo	sly-miR6027-3p	0,0
MSTRG.37256.1	novo	sly-miR6027-5p	0,0
MSTRG.30956.1	novo	sly-miR9474-3p	0,0
MSTRG.6806.1	novo	sly-miR167a	0,5
MSTRG.6806.1	novo	sly-miR167b-3p	1,0

Fonte: Elaborado pelos autores deste artigo.

Uma das formas de inferir as possíveis funções executadas pelos lncRNAs é por meio das funções das moléculas que interagem com eles, como os miRNAs. Na **Figura 4** podemos observar as funções descritas, na literatura, para os miRNAs preditos pelo psRNATarget.

Figura 4: Funções descritas na literatura para as famílias de miRNAs preditas pelo psRNATarget.

LncRNA	miRNA	Caracterização
MSTRG.14093.1 MSTRG.29733.2 MSTRG.29733.1	conhecido novo miR156	 Regulação da fase vegetativa; formação de raiz adventícia <i>M. xiaojinensis</i> ; controle de processos no fruto em tomate.
MSTRG.6806.1	novo miR167	Resposta a axinas e desenvolvimento floral em tomate.
MSTRG.24581.1 MSTRG.37256.1	novo miR172 miR6027	 Identidade e número de órgãos florais em tomate; desenvolvimento e florescimento de plantas. Clivagem de genes de resistência em solanáceas;
MSTRG.30956.1 MSTRG.22083.2 MSTRG.22083.4 MSTRG.22083.5	novo conhecido miR9474 miR9470	Resposta a seca.  Relacionado a vias de etileno e carotenóides em tomate.

Fonte: Elaborada pelos autores deste artigo.

Pelas funções dos miRNAs que clivam esses lncRNAs, é possível prever os processos nos quais estão envolvidos os lncRNAs. Mir156 está envolvido na regulação de mudança da fase vegetativa em vegetais superiores, através da interação com seus alvos, os genes-alvo SBP/SPL (JIA et al. 20117). Mir167, através da regulação de seus genes-alvo (ARF6 e ARF8), induz a defesa dependente de ácido salicílico contra *Pseudomonas syringae* em *Arabidopsis* (CARUANA; DHAR; RAINA, 2020). Mir172 é uma família de microRNAs conservada em espermatófitos, e está envolvido na transição floral em *Arabidopsis* (ZHANG; CHEN, 2021). Mir6027 foi identificado

somente em algumas Solanaceae, e regula genes de resistência (SEO et al. 20018). Mir9474 está envolvido na via de transdução de sinais hormonais, e tem como alvos uma proteína TIFY e uma fosfatase 2c (PP2C) (DONG et al. 2020). Mir9470, em tomate, tem como alvo ERF021, e está envolvido com a via de regulação do etileno (MA et al. 2020).

Para as análises posteriores, foram selecionados os lncRNAs que apresentaram maior diferença de expressão entre dois tratamentos e os lncRNAs identificados como alvos de miRNAs (**Quadro 3**). A predição da localização subcelular apontou, principalmente, o citoplasma, com apenas dois lncRNAs (MSTRG.1361.1 e MSTRG.40151.1) com localização predita para o núcleo.

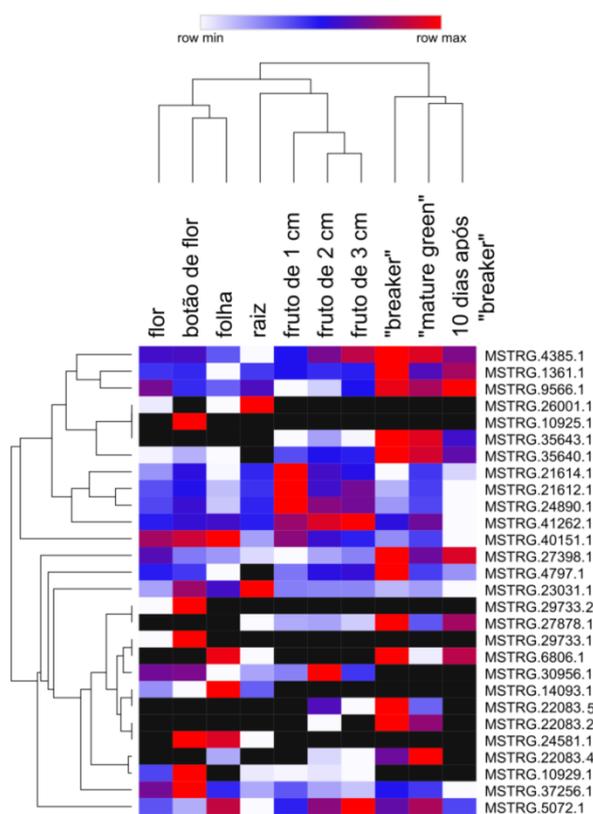
Quadro 3: LncRNAs com maior diferença de expressão entre dois tratamentos e lncRNAs identificados como alvos de miRNAs.

LncRNAs preditos como alvos de miRNAs				
lncRNA	Tamanho (nt)	Anotação NCBI	Cromossomo	Localização
MSTRG.14093.1	963	ncRNA	SL4.0ch03	Citoplasma
MSTRG.22083.2	1618	ncRNA	SL4.0ch06	Citoplasma
MSTRG.22083.4	1325	ncRNA	SL4.0ch06	Citoplasma
MSTRG.22083.5	1046	ncRNA	SL4.0ch06	Citoplasma
MSTRG.29733.2	409	sem anotação	SL4.0ch08	Citoplasma
MSTRG.29733.1	527	sem anotação	SL4.0ch08	Citoplasma
MSTRG.6806.1	320	sem anotação	SL4.0ch02	Citoplasma
MSTRG.24581.1	457	ncRNA	SL4.0ch06	Citoplasma
MSTRG.37256.1	621	sem anotação	SL4.0ch10	Citoplasma
MSTRG.30956.1	403	sem anotação	SL4.0ch08	Citoplasma
LncRNAs com a maior diferença de expressão				
MSTRG.1361.1	619	mRNA	SL4.0ch01	Núcleo
MSTRG.4385.1	440	ncRNA	SL4.0ch01	Citoplasma
MSTRG.4797.1	519	sem anotação	SL4.0ch01	Citoplasma
MSTRG.5072.1	446	sem anotação	SL4.0ch01	Citoplasma
MSTRG.9566.1	921	mRNA	SL4.0ch02	Citoplasma
MSTRG.10925.1	478	ncRNA	SL4.0ch03	Citoplasma
MSTRG.10929.1	477	ncRNA	SL4.0ch03	Citoplasma
MSTRG.21612.1	368	ncRNA	SL4.0ch06	Citoplasma
MSTRG.21614.1	399	ncRNA	SL4.0ch06	Citoplasma
MSTRG.23031.1	548	ncRNA	SL4.0ch06	Citoplasma
MSTRG.24890.1	385	long non-coding RNA	SL4.0ch06	Citoplasma
MSTRG.26001.1	429	ncRNA	SL4.0ch07	Citoplasma
MSTRG.27398.1	548	ncRNA	SL4.0ch07	Citoplasma
MSTRG.27878.1	508	sem anotação	SL4.0ch07	Citoplasma
MSTRG.35640.1	471	sem anotação	SL4.0ch10	Citoplasma
MSTRG.35643.1	479	sem anotação	SL4.0ch10	Citoplasma
MSTRG.40151.1	335	ncRNA	SL4.0ch11	Núcleo
MSTRG.41262.1	394	sem anotação	SL4.0ch12	Citoplasma

Fonte: Elaborado pelos autores deste artigo.

Os lncRNAs que apresentaram maior diferença de expressão entre as comparações e os alvos de miRNAs foram selecionados e utilizados para a confecção de um *heatmap* de expressão, com os valores convertidos para log₂(TPM) (**Figura 5**):

Figura 5: Os 28 lncRNAs selecionados para análises de expressão. Foram selecionados os lncRNAs identificados como alvos de miRNAs e os lncRNAs com maior diferença de expressão entre dois tratamentos.

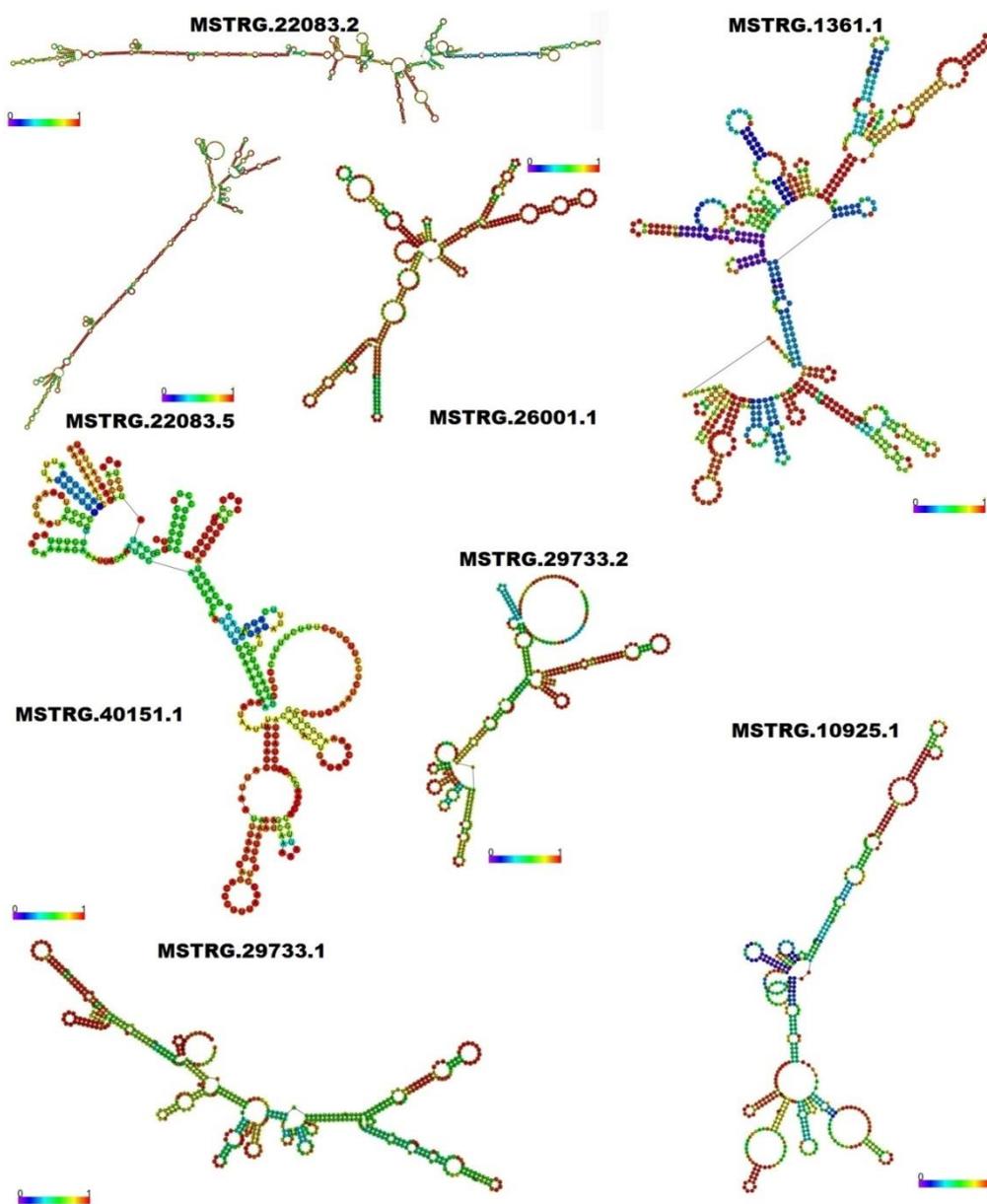


Fonte: Elaborado pelos autores.

De forma geral, percebemos que os lncRNAs possuem maior expressão apenas em uma ou poucas comparações, o que condiz com os dados da literatura, que apontam para uma alta especificidade de lncRNAs, quando se considera diferentes tecidos, tratamentos e estádios do desenvolvimento vegetal (WANG; CHEKANOVA, 2017). MSTRG10925.1, por exemplo, possui expressão apenas no botão floral, e MSTRG24581.1 tem alta expressão apenas em botão de flor e em folha.

Alguns lncRNAs foram selecionados para modelagem (**Figura 6**).

Figura 6: Estrutura secundária dos lncRNAs selecionados.

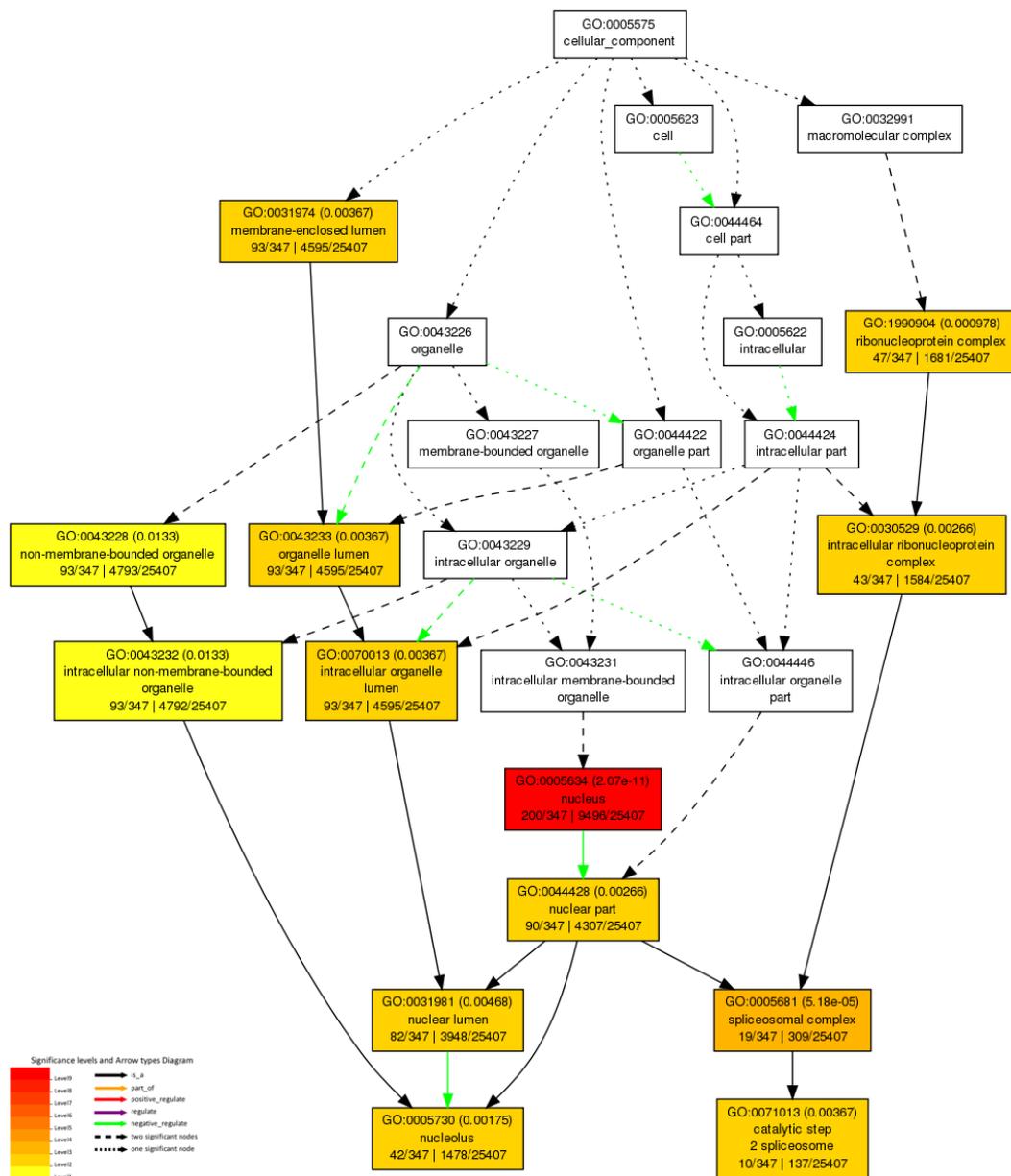


Fonte: Elaborado pelos autores deste artigo.

A predição da estrutura secundária dos lncRNAs revelou que os mesmos são consideravelmente diferentes entre si. Dependendo da função exercida pelos lncRNAs, a estrutura secundária pode desempenhar um papel significativo. Apesar da baixa conservação em nível de nucleotídeos, muitos ncRNAs possuem alta conservação da estrutura secundária (SANBONMATSU, 2016).

Foi selecionado o lncRNA MSTRG.24890.1 para a análise de GO (**Figura 7**).

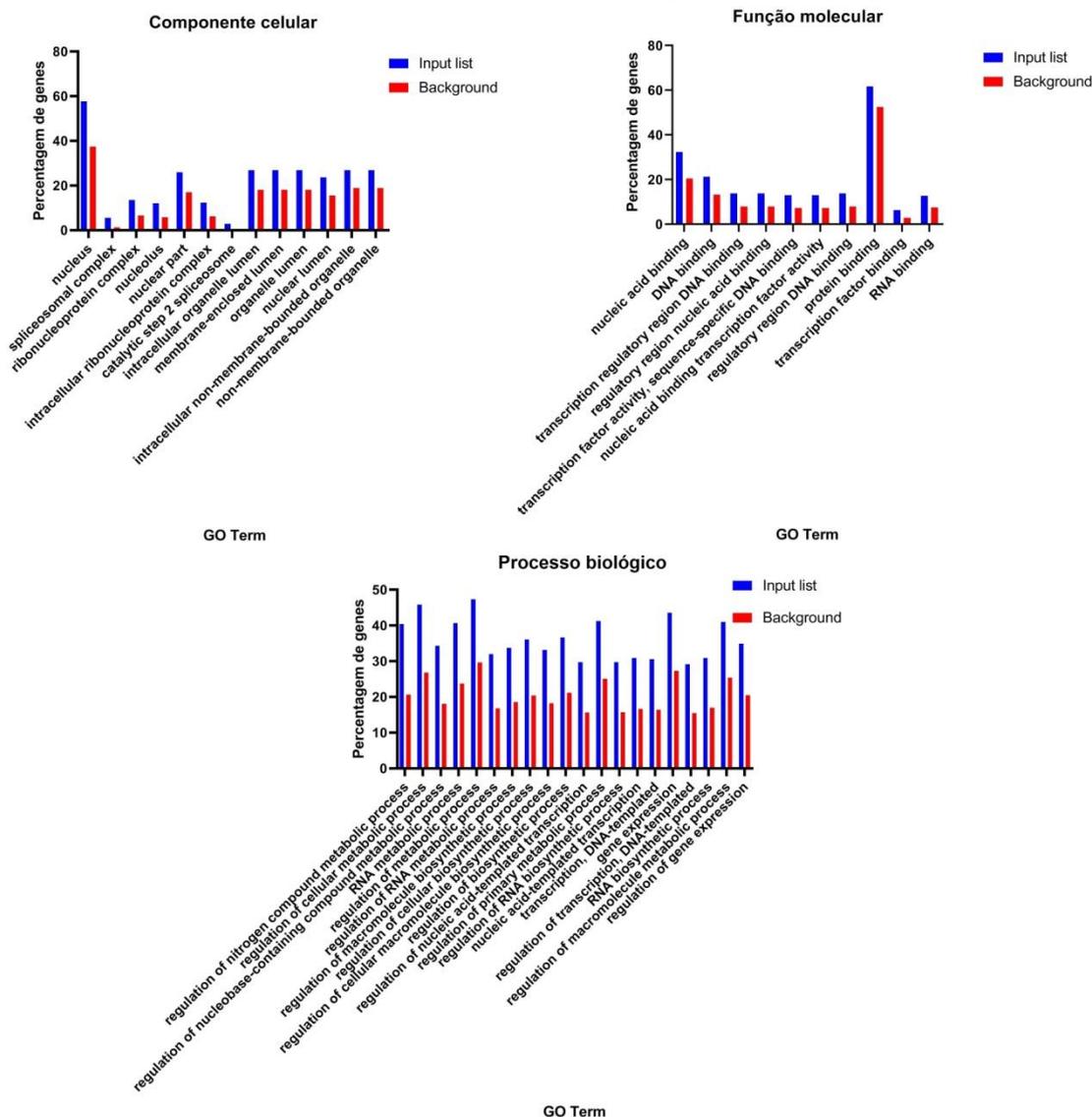
Figura 7: Componente celular dos mRNAs preditos como alvos do lncRNA MSTRG.24890.1.



Fonte: Imagens produzidas pelos autores com uso do agriGO v2.0.

MSTRG.24890.1 possui alta expressão em frutos (sendo mais alta em fruto de 1 cm), e seus alvos foram preditos por meio do programa IntaRNA. A análise de GO dos mRNAs preditos como alvos do lncRNA MSTRG.24890.1 revelou enriquecimento para o núcleo como componente celular (**Figura 7**) e ligação a ácido nucléico como função molecular (dados não mostrados). Assim, a função do lncRNA pode ser indiretamente inferida a partir das funções de suas proteínas alvos preditas por GO (ZHANG; ZOU; DENG, 2017). As anotações dos alvos enriquecidos estão representadas na **Figura 8**.

Figura 8: Análise de enriquecimento de GO dos mRNAs preditos como alvos dos lncRNAs.



Fonte: Gráficos produzidos pelos autores a partir dos dados gerados pelo agriGO.

Os processos biológicos com enriquecimento pela análise de GO foram diversos, e para a construção do gráfico selecionamos apenas os 20 primeiros processos listados, com destaque para: regulação da expressão gênica; processos metabólicos de RNA; regulação de processos metabólicos de nitrogênio e regulação de processos metabólicos celulares. Os processos biológicos nos quais estão envolvidos os alvos do lncRNA MSTRG.24890.1 (e o próprio lncRNA) apontam para um envolvimento deste lncRNA com ácidos nucleicos, o que também se reflete no componente celular principal apontado pelo GO (núcleo) e com as funções moleculares (ligação a proteína e ligação a ácido nucleico). Isso sugere que este lncRNA pode exercer sua função interagindo com DNA ou RNA.

4 CONCLUSÕES

A identificação de lncRNAs envolvidos em diferentes processos de desenvolvimento em tomate constitui um importante passo para uma maior compreensão do papel desempenhado por essas importantes moléculas não-codificantes, que têm sido descritas como elementos-chave da regulação dos mais diferentes processos. No caso específico do tomate, identificar lncRNAs envolvidos em processos comercialmente importantes, como o amadurecimento do fruto, pode auxiliar pesquisadores no melhoramento deste cultivar.

Neste trabalho, por meio de análises de bioinformática, identificamos lncRNAs diferencialmente expressos entre diferentes tecidos e estádios de desenvolvimento de fruto de tomate. Foram identificados diversos miRNAs que podem clivar os lncRNAs. A análise de expressão de lncRNAs selecionados sugere uma especificidade de tecido e estágio de desenvolvimento, com alguns lncRNAs expressos apenas em determinados tecidos. A obtenção da estrutura secundária dos lncRNAs revelou uma alta divergência entre os mesmos, que pode estar associada às funções exercidas. A predição dos alvos do lncRNA MSTRG.24890.1 sugere sua interação com material genético, DNA ou RNA, sendo o núcleo apontado como o compartimento onde estão localizados os alvos de MSTRG.24890.1.

Esperamos, com este trabalho, contribuir para uma maior compreensão acerca dos papéis desempenhados por lncRNAs em tomate, fornecendo subsídios para futuras pesquisas e obtenção de cultivares mais produtivos e resistentes a diferentes tipos de estresse.

REFERÊNCIAS

ANDREWS, S. FastQC: a quality control tool for high throughput sequence data, 2010. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Acesso em: 03 jul. 2022.

BOLGER, Anthony M.; LOHSE, Marc; USADEL, Bjoern. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, vol. 30, n. 15, p. 2114–2120, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4103590/pdf/btu170.pdf>. Acesso em: 03 jul. 2022.

BRAY, N., PIMENTEL, H., MELSTED, P. et al. Near-optimal probabilistic RNA-seq quantification. **Nat Biotechnol**, vol. 34, p. 525–527, 2016. Disponível em: <https://www.nature.com/articles/nbt.3519>. Acesso em: 03 jul. 2022.

CAMARGO, A. P.; SOURKOV, V.; PEREIRA, G. A. G.; CARAZZOLLE, M. F. RNAsamba: neural network-based assessment of the protein-coding potential of RNA sequences. **NAR Genomics and Bioinformatics**, vol. 2, n. 1, 2020. Disponível em: <https://academic.oup.com/nargab/article/2/1/lqz024/5701461>. Acesso em: 04 jul. 2022.

CARUANA, Julie C.; DHAR, Nikhilesh; RAINA, Ramesh. Overexpression of Arabidopsis microRNA167 induces salicylic acid-dependent defense against *Pseudomonas syringae* through the regulation of its targets ARF6 and ARF8. **Plant Direct.**, p. 1–16, 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7510475/pdf/PLD3-4-e00270.pdf>. Acesso em: 04 jul. 2022.

CARVALHO, Carla Roberta Ferraz; PONCIANO, Niraldo José; SOUZA, Paulo Marcelo de; SOUZA, Cláudio Luiz Melo de; SOUSA, Elias Fernandes de. Viabilidade econômica e de risco da produção de tomate no município de Cambuci/RJ, Brasil. **Ciência Rural**, v.44, n.12, p. 2293-2299, 2014. Disponível em: <https://www.scielo.br/j/cr/a/pFZxWYby8KkgjFcypsgP5Tk/?format=pdf&lang=pt>. Acesso em: 02 jul. 2022.

COELHO, Raissa Gomes; OLIVEIRA, Francisca Djanira Fernandes de; SOUZA, Edivaldo Bezerra de; AZEVEDO, José Marlo Araújo de; LIMA, Maurifran Oliveira. Desenvolvimento e características produtivas de tomate do tipo cereja em diferentes compostos orgânicos. **Revista Spacios**, vol. 39, n. 26, 2018. Disponível em: <http://www.revistaespacios.com/a18v39n26/a18v39n26p32.pdf>. Acesso em: 02 jul. 2022.

CONAB, **Tomate: Análise dos Indicadores da Produção e Comercialização no Mercado Mundial, Brasileiro e Catarinense**, vol. 21, 2019. Disponível em: https://www.conab.gov.br/institucional/publicacoes/compendio-de-estudos-da-conab/item/download/29586_4fe6dd2c9c6d1fa5e1cbc5f82061717d. Acesso em: 02 jul. 2022.

COSTA, Fabiano Guimarães; FILHO, José Vicente Caixeta. Análise das perdas na comercialização de tomate: um estudo de caso. **Revista Informações Econômicas**, v.26, n.12, 1996. Disponível em: <http://www.iea.sp.gov.br/ftp/iea/ie/1996/tec1-1296.pdf>. Acesso em: 02 jul. 2022.

CUI, Jun; JIANG, Ning; HOU, Xinxin; WU, Sihan; ZHANG, Qiang; MENG, Jun; LUAN, Yushi. Genome-Wide Identification of lncRNAs and Analysis of ceRNA Networks During Tomato Resistance to *Phytophthora infestans*. **Phytopathology**, vol. 110, p. 456-464, 2020. Disponível em: <https://apsjournals.apsnet.org/doi/epdf/10.1094/PHYTO-04-19-0137-R>. Acesso em: 03 jul. 2022.

DAI, Xinbin; ZHAO, Patrick Xuechun. psRNATarget: a plant small RNA target analysis server. **Nucleic Acids Research**, vol. 39, 2011. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3125753/pdf/gkr319.pdf>. Acesso em: 04 jul. 2022.

DONG, Fei; WANG, Chuanzeng; DONG, Yuhui; HAO, Shuqin; WANG, Lixia; SUN, Xiudong; LIU, Shiqi. Differential expression of microRNAs in tomato leaves treated with different light qualities. **BMC Genomics**, vol. 21, n. 37, 2020. Disponível em: <https://bmcbgenomics.biomedcentral.com/track/pdf/10.1186/s12864-019-6440-4.pdf>. Acesso em: 04 jul. 2022.

FERREIRA, Sila Mary Rodrigues. **Características de qualidade do tomate de mesa (*Lycopersicon esculentum* Mill.) cultivado nos sistemas convencional e orgânico comercializado na região metropolitana de Curitiba**. Tese (Doutorado em Tecnologia de Alimentos) – Universidade Federal do Paraná, Curitiba, p. 249. 2004.

GALLART, Andreu Paytuví; PULIDO, Antonio Hermoso; LAGRÁN, Irantzu Anzar Martínez de; SANSEVERINO, Walter; CIGLIANO, Riccardo Aiese. GREENC: a Wiki-based database of plant lncRNAs. **Nucleic Acids Research**, vol. 44, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702861/pdf/gkv1215.pdf>. Acesso em: 04 jul. 2022

GOODSTEIN, David M.; SHU, Shengqiang; HOWSON, Russell; NEUPANE, Rochak; HAYES, Richard D.; FAZO, Joni; MITROS, Therese; DIRKS, William; HELLSTEN, Uffe; PUTNAM, Nicholas; ROKHSAR, Daniel S. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, vol. 40, 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245001/pdf/gkr944.pdf>. Acesso em: 03 jul. 2022.

GRIFFITHS-JONES, Sam; BATEMAN, Alex; MARSHALL, Mhairi; KHANNA, Ajay; EDDY, Sean R. Rfam: an RNA family database. **Nucleic Acids Research**, vol. 31, n. 1, 2003. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC165453/pdf/gkg006.pdf>. Acesso em: 03 jul. 2022.

GRUBER, Andreas R.; LORENZ, Ronny; BERNHART, Stephan H.; NEUBÖCK, Richard; HOFACKER, Ivo L. The Vienna RNA Websuite. **Nucleic Acids Research**, 2008, Vol. 36. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2447809/pdf/gkn188.pdf>. Acesso em: 04 jul. 2022.

GUO, Jin-Cheng; FANG, Shuang-Sang; WU, Yang; ZHANG, Jian-Hua; CHEN, Yang; LIU, Jing; WU, Bo; WU, Jia-Rui; LI, En-Min; XU, Li-Yan; SUN, Liang; ZHAO, Yi. CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition. **Nucleic Acids Research**, vol. 47, 2019. Disponível em: <https://academic.oup.com/nar/article/47/W1/W516/5506859?login=false>. Acesso em: 04 jul. 2022.

JIA, Xiao Lin; CHEN, Ya Kun; XU, Xiao Zhao et al. miR156 switches on vegetative phase change under the regulation of redox signals in apple seedlings. **Scientific Reports**, vol. 7, artigo n. 14223, 2017. Disponível em: <https://www.nature.com/articles/s41598-017-14671-8.pdf>. Acesso em: 04 jul. 2022.

JIANG, Ning; CUI, Jun; HOU, Xinxin; YANG, Guanglei; XIAO, Yu; HAN, Lu; MENG, Jun; LUAN, Yushi. Sl-lncRNA15492 interacts with Sl-miR482a and affects *Solanum lycopersicum* immunity against *Phytophthora infestans*. **The Plant Journal**, vol. 103, p. 1561–1574, 2020.

Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/tpj.14847>. Acesso em: 03 jul. 2022.

KANG, Yu-Jian; YANG, De-Chang; KONG, Lei; HOU, Mei; MENG, Yu-Qi; WEI, Liping; GAO, Ge. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. **Nucleic Acids Research**, vol. 45, 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5793834/pdf/gkx428.pdf>. Acesso em: 04 jul. 2022.

KONG, Lei; ZHANG, Yong; YE, Zhi-Qiang; LIU, Xiao-Qiao; ZHAO, Shu-Qi; WEI, Liping; GAO, Ge. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. **Nucleic Acids Research**, vol. 35, 2007. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1933232/pdf/gkm391.pdf>. Acesso em: 03 jul. 2022.

LI, Ning; WANG, Zhongyu; WANG, Baike; WANG, Juan; XU, Ruiqiang; YANG, Tao; HUANG, Shaoyong; YU, Qinghui; WANG, Huan; GAO, Jie. Identification and Characterization of Long Non-Coding RNA in Tomato Roots under Salt Stress. **bioRxiv preprint**. Disponível em: <https://www.biorxiv.org/content/10.1101/2021.11.12.468377v1.full.pdf>. Acesso em: 03 jul. 2022.

MA, Lili; MU, Jianlou; GRIERSON, Donald et al. Noncoding RNAs: functional regulatory factors in tomato fruit ripening. **Theoretical and Applied Genetics**, 2020. Disponível em: <https://ur.booksc.me/dl/81546234/12b0b4>. Acesso em: 04 jul. 2022.

MANN M, WRIGHT PR, BACKOFEN R. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. **Nucleic Acids Res.**, vol. 45, p. W435–W439, 2017. Disponível em: http://www.bioinf.uni-freiburg.de/Publications/Mann_Wright_Backofen-IntaR_enhan_and-NAR2017.pdf. Acesso em: 01 jul. 2022.

NIAZI, FARSHAD; VALADKHAN, SABA. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 39 UTRs. **RNA**, vol. 18, p. 825–843, 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3312569/pdf/825.pdf>. Acesso em: 05 jul. 2022.

PEREIRA-CARVALHO, Rita de Cássia; TOBAR, Leidy Lorena Mendoza; DIANESE, Erico de Campos; FONSECA, Maria Esther de Noronha; BOITEUX, Leonardo Silva. Melhoramento genético do tomateiro para resistência a doenças de etiologia viral: avanços e perspectivas. **Revisão Anual de Patologia de Plantas (RAPP)**, vol. 22, 2014. Disponível em: <https://repositorio.bc.ufg.br/bitstream/ri/14209/5/Artigo%20-%20Rita%20de%20C%20C3%A1ssia%20Pereira-Carvalho%20-%202014.pdf>. Acesso em: 02 jul. 2022.

PERTEA, Geo; PERTEA, Mihaela. GFF Utilities: GffRead and GffCompare. **F1000Research**, vol. 9, n.304, 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7222033/pdf/f1000research-9-26865.pdf>. Acesso em: 03 jul. 2022.

PERTEA, M., PERTEA, G., ANTONESCU, C. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. **Nat Biotechnol**, vol. 33, p. 290–295, 2015. Disponível em: <https://www.nature.com/articles/nbt.3122>. Acesso em: 03 jul. 2022.

RITU; GUPTA, Sagar; SHARMA, Nitesh Kumar; SHANKAR, Ravi. DeepPlnc: Discovering plant lncRNAs through multimodal deep learning on sequential data. **bioRxiv preprint**. Disponível em: <https://www.biorxiv.org/content/10.1101/2021.12.10.472074v1.full.pdf>. Acesso em: 04 jul. 2022.

SANBONMATSU, Karissa. Towards structural classification of long non-coding RNAs. **Biochim Biophys Acta**, art. n. 1859, vol. 1, p. 41–45, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6827331/pdf/nihms-1046098.pdf>. Acesso em: 04 jul. 2022.

SEO, Eunyoung; KIM, Taewook; PARK, June Hyun; YEOM, Seon-In; KIM, Seungill; SEO, Min-Ki; SHIN, Chanseok; CHOI, Doil. Genome-wide comparative analysis in Solanaceous species reveals evolution of microRNAs targeting defense genes in *Capsicum* spp. **DNA Research**, vol. 25 n.6, p. 561–575, 2018. Disponível em: <https://academic.oup.com/dnaresearch/article/25/6/561/5060358>. Acesso em: 04 jul. 2022.

SZCZESŃIAK, Michal W.; ROSIKIEWICZ, Wojciech; MAKALOWSKA, Izabela. CANTATAdb: A Collection of Plant Long Non-Coding RNAs. **Plant Cell Physiol.**, vol. 57, n. 1, p. e8(1–7), 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4722178/pdf/pcv201.pdf>. Acesso em: 04 jul. 2022.

TIAN, Tian; LIU, Yue; YAN, Hengyu; YOU, Qi; YI, Xin; DU, Zhou; XU, Wenying; SU, Zhen. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. **Nucleic Acids Research**, vol. 45, n. W1, 32017, p. W122–W129, 2017. Disponível em: <https://academic.oup.com/nar/article/45/W1/W122/3796337>. Acesso em: 04 jul. 2022.

WANG, Hsiao-Lin V.; CHEKANOVA, Julia A. Long Noncoding RNAs in Plants. **Adv Exp Med Biol.**, vol. 1008, p. 133–154, 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6689229/pdf/nihms-1043265.pdf>. Acesso em: 03 jul. 2022.

WANG, Jinyan; YANG, Yuwen; JIN, Lamei; LING, Xitie; LIU, Tingli; CHEN, Tianzi; JI, Yinghua; YU, Wengui; ZHANG, Baolong. Re-analysis of long non-coding RNAs and prediction of circRNAs reveal their novel roles in susceptible tomato following TYLCV infection. **BMC Plant Biology**, vol. 18, n. 104, 2018b. Disponível em: <https://bmcpantbiol.biomedcentral.com/track/pdf/10.1186/s12870-018-1332-3.pdf>. Acesso em: 03 jul. 2022.

WANG, Jinyan; YU, Wengui; YANG, Yuwen; LI, Xiao; CHEN, Tianzi; LIU, Tingli; MA, Na; YANG, Xu; LIU, Renyi; ZHANG, Baolong. Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. **Scientific Reports**, vol. 5, n. 16946, 2015. Disponível em: <https://www.nature.com/articles/srep16946.pdf>. Acesso em: 03 jul. 2022.

WANG, Minghui; ZHAO, Weihua; GAO, Lei; ZHAO, Lingxia. Genome-wide profiling of long non-coding RNAs from tomato and a comparison with mRNAs associated with the regulation of fruit ripening. **BMC Plant Biology**, vol. 18, n.75, 2018a. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5935960/pdf/12870_2018_Article_1300.pdf. Acesso em: 03 jul. 2022.

ZENG, Min; WU, Yifan; LU, Chengqian; ZHANG, Fuhao; WU, Fang-Xiang; LI, Min. DeepLncLoc: a deep learning framework for long non-coding RNA subcellular localization prediction based on subsequence embedding. **Briefings in Bioinformatics**, vol. 23, n. 1, 2022. Disponível em: <https://academic.oup.com/bib/article-abstract/23/1/bbab360/6366323?redirectedFrom=fulltext&login=false>. Acesso em: 04 jul. 2022.

ZHANG, Bailong; CHEN, Xuemei. Secrets of the MIR172 family in plant development and flowering unveiled. Secrets of the MIR172 family in plant development and flowering unveiled. **PLoS Biol**, vol. 19, n.2, p. e3001099. Disponível em: <https://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.3001099&type=printable>. Acesso em: 04 jul. 2022.

ZHANG, Jingpu; ZOU, Shuai; DENG, Lei. BiRWLGO: A global network-based strategy for lncRNA function annotation using bi-random walk. **IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**, pp. 50-55, 2017. Disponível em: <https://ieeexplore.ieee.org/document/8217623>. Acesso em: 04 jul. 2022.