

ANÁLISE COMPARATIVA DE PREDITORES DE RNA CIRCULAR CONTRA BANCOS DE mRNA-Seq DE *Fragaria* x *ananassa*

HUGO CARLOS BOLZON GONZALEZ¹; CHRISTIAN SANCHEZ²; TATIANE JÉSSICA SIEBENEICHLER³; VANESSA GALLI⁴

¹Universidade Federal de Pelotas – hugocarlos.bg@gmail.com

²Universidade Federal de Pelotas – christian.kun@gmail.com

³Universidade Federal de Pelotas – tatianejs1@hotmail.com

⁴Universidade Federal de Pelotas – vane.galli@yahoo.com.br

1. INTRODUÇÃO

Os RNAs circulares, ou circRNAs, são uma classe distinta de RNAs não codificantes (ncRNAs) que se originam de eventos de *backsplicing* durante a maturação do mRNA (CHEN; YANG, 2015). Eles possuem uma estrutura molecular única, formando uma alça fechada covalentemente sem terminações 5' ou 3', e sem cauda poliadenilada (JECK et al., 2013). Esta estrutura circular confere aos circRNAs uma estabilidade superior às moléculas de RNA linear, tornando-os resistentes à ação de exonucleases (PETKOVIC & MULLER, 2015).

A biogênese dos circRNAs ocorre durante o processamento do pré-mRNA. O RNA nascente pode ser direcionado para vias alternativas que facilitam o *backsplicing*. Este processo é mediado por emparelhamento de bases entre elementos repetitivos invertidos localizados em introns *upstream* e *downstream*, ou por dimerização de *RNA-binding proteins* (RBPs) que se ligam com *motifs* específicos nos introns flanqueadores (P. ZHANG et al., 2020).

Diversas funções biológicas têm sido atribuídas aos RNAs circulares, dentre as quais a atuação como esponjas de miRNAs, controlando indiretamente a expressão gênica. Eles se ligam a miRNAs e impedem que eles se liguem aos seus alvos gênicos, atuando como reguladores negativos da expressão gênica (PANDA; GOROSPE, 2018). Ainda, estão envolvidos na regulação da resposta a estresses bióticos e abióticos em plantas, bem como na maturação de frutos, como demonstrado em trigo (YUEXIA WANG et al., 2017) e tomate (TAN et al., 2017), respectivamente.

A identificação de circRNAs é feita através de ferramentas computacionais que empregam algoritmos baseados em *back-splicing junction* (BSJ). Essas ferramentas apresentam diferenças nos algoritmos de predição, bem como nos alinhamentos e mapeamentos, incorporando ou não funcionalidades adicionais como a predição de interações entre circRNAs e miRNA (CHEN et al., 2021).

Os preditores desempenham um papel crucial na identificação precisa dos circRNAs. A maioria dos preditores utiliza uma etapa de mapeamento das BSJ. No entanto, existem abordagens alternativas, como as baseadas em machine learning (G. ZHANG et al., 2020) e ferramentas baseadas em K-mers (JIANG et al., 2021). Ao escolher um preditor, é importante considerar sua especificidade para reduzir a incidência de falsos positivos. Devido à divergência entre os preditores, é benéfico utilizar mais de um para obter uma análise mais precisa e completa.

O objetivo do trabalho é realizar uma comparação direta entre os preditores find_circ e CIRI2 utilizando o mesmo conjunto de dados de RNA-seq. Ao fazer isso, o estudo espera avaliar a utilização das ferramentas de predição, contra

bases de dados de plantas, fornecendo *insights* sobre a eficácia desses dois preditores na identificação de circRNAs.

2. METODOLOGIA

Os dados de RNA-seq utilizados foram obtidos do *Genbank Sequence Read Archive* (SRA), acessível pelo código PRJNA338879 e publicado no GEO (*Gene Expression Omnibus*) com o número GSE85572. Este conjunto inclui 14 bibliotecas de mRNA-seq, derivadas a partir de amostras de frutos de morango cultivados com ou sem aplicação de ABA ou seu inibidor (NDGA), em três estádios diferentes (dias 0, 5 e 8), sendo o dia zero definido como duas semanas após a antese (fase de abertura da flor), e os dias 5 e 8 como 5 e 8 dias após o dia 0, respectivamente. O sequenciamento foi realizado utilizando a plataforma Illumina HiSeq 4000. A escolha deste conjunto de dados se deu por conter diferentes estresses e tratamentos, aumentando assim a probabilidade dos circRNAs estarem expressos e sequenciados no sequenciamento do mRNA-seq.

No entanto, para esta análise preliminar comparativa entre os preditores, optou-se por processar uma fração desses dados, devido ao tamanho das bibliotecas. Os bancos priorizados foram:

- Fruto de morango tratado com água duas semanas após a antese, definido como dia 0 - SRR4030803
- Fruto de morango tratado com água, amostras coletadas 5 dias após o tratamento - SRR4030805
- Fruto de morango tratado com ABA, amostras coletadas 5 dias após o tratamento - SRR4030809
- Fruto de morango tratado com NDGA, amostras coletadas 5 dias após o tratamento - SRR4030813

O genoma de referência utilizado foi o *Fragaria x ananassa Genome v1.0* (FAN_r1.1.fasta), que representa quase a totalidade do genoma da espécie de morango octaplóide *Fragaria ananassa* Camarosa. Este genoma foi obtido do banco de dados genômicos de Rosaceas (GDR, *Genome Database for Rosaceae*). Além disso, o genoma anotado *Fragaria ananassa Genome v1.0* (FAN_r1.1.gff3) também foi utilizado, igualmente obtido do mesmo banco.

A fim de comparar diferentes preditores de circRNA, optamos por utilizar as ferramentas find_circ e CIRI2. O alinhamento e mapeamento prévio das leituras para o find_circ foram realizados com a ferramenta Bowtie2, enquanto para o CIRI2 utilizou-se a ferramenta BWA, com o algoritmo BWA MEM. Em ambas as análises, utilizamos o mesmo genoma de referência e genoma anotado, permitindo uma comparação direta entre os dois preditores.

Devido à quantidade significativa de dados originados pelos preditores exigiu-se a execução de uma etapa de filtragem, que foi realizada considerando sequências preditas que possuíam mais de 10 leituras e que são identificadas como 'circulares'. Posteriormente, sequências com mesma localização cromossômica e tamanho foram unificadas. A lista final de circRNAs preditos foi então analisada para detalhar sua composição.

3. RESULTADOS E DISCUSSÃO

No estudo, foram utilizados bancos de RNA-seq de morangos (*Fragaria x ananassa*) submetidos a tratamentos exógenos de ABA, NDGA e controle. Após o alinhamento e mapeamento com a ferramenta Bowtie2, o preditor find_circ identificou possíveis candidatos a circRNAs. A filtragem resultou em 31 circRNAs, com um comprimento médio de 977 pb. Destes circRNAs, 3 (9,7%) apresentavam um comprimento entre 100 e 200 pb, enquanto que 8 (25,8%) apresentavam um comprimento entre 200 e 300 pb. Além disso, foi observado que 2 circRNAs (6,4%) apresentavam um comprimento entre 300 e 400 pb e que outros 18 circRNAs (58,1%) apresentavam um comprimento superior a 400 pb.

Da mesma forma, após o alinhamento e mapeamento com a ferramenta BWA, o preditor CIRI2 também foi utilizado para identificar possíveis candidatos à circRNAs. A filtragem resultou em 36 circRNAs, com um comprimento médio de 1532 pb. Destes circRNAs, 0 (0,0%) apresentavam um comprimento entre 100 e 200 pb, enquanto que 2 (5,5%) apresentavam um comprimento entre 200 e 300 pb. Além disso, foi observado que 0 circRNAs (0,0%) apresentavam um comprimento entre 300 e 400 pb e que outros 34 circRNAs (94,5%) apresentavam um comprimento superior a 400 pb.

Os resultados estão demonstrados na tabela 1 a seguir.

Tabela 1. Resultado dos preditores de circRNA agrupados por tamanho de fragmento

Faixa de tamanho (pb)	Preditores de circRNA	
	find_circ	CIRI2
100-200	3 (9,7%)	0 (0,0%)
200-300	8 (25,8%)	2 (5,5%)
300-400	2 (6,4%)	0 (0,0%)
>400	18 (58,1%)	34 (94,5%)

4. CONCLUSÕES

O presente estudo realizou uma análise comparativa entre diferentes preditores de RNA circular, demonstrando divergência nos resultados, independente da utilização dos mesmos bancos de dados base. Tais divergências corroboram a percepção de que a utilização de diferentes ferramentas de predição e identificação de circRNAs é uma prática benéfica, que pode ampliar e enriquecer o levantamento destes RNAs em bancos de dados de sequenciamento biológico.

Ainda, os contrastes apresentados pelos preditores sugere a necessidade de ampliar a base de dados de análise, processando mais sequenciamentos e comparando as divergências e convergências entre as sequências de circRNA preditas, de modo a produzir um levantamento consensual e mais fidedigno das predições.

As análises *in silico* para prospecção de RNAs circulares se apresenta como uma metodologia ascendente e robusta, na identificação e elucidação desta classe de RNAs, e a integração de diferentes ferramentas tende a ampliar ainda mais este conhecimento gerado.

5. REFERÊNCIAS BIBLIOGRÁFICAS

CHEN, Liang et al. The bioinformatics toolbox for circRNA discovery and analysis. **Briefings in bioinformatics**, v. 22, n. 2, p. 1706-1728, 2021.

CHEN, Ling-Ling; YANG, Li. Regulation of circRNA biogenesis. **RNA biology**, v. 12, n. 4, p. 381-388, 2015.

JECK, William R. et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. **Rna**, v. 19, n. 2, p. 141-157, 2013.

JIANG, Jyun-Yu et al. JEDI: circular RNA prediction based on junction encoders and deep interaction among splice sites. **Bioinformatics**, v. 37, n. Supplement_1, p. i289-i298, 2021.

PANDA, Amaresh C.; GOROSPE, Myriam. Detection and analysis of circular RNAs by RT-PCR. **Bio-protocol**, v. 8, n. 6, p. e2775-e2775, 2018.

PETKOVIC, Sonja; MÜLLER, Sabine. RNA circularization strategies in vivo and in vitro. **Nucleic acids research**, v. 43, n. 4, p. 2454-2465, 2015.

TAN, Jinjuan et al. Identification and functional characterization of tomato circRNAs derived from genes involved in fruit pigment accumulation. **Scientific Reports**, v. 7, n. 1, p. 8594, 2017.

WANG, Yuexia et al. Identification of circular RNAs and their targets in leaves of *Triticum aestivum* L. under dehydration stress. **Frontiers in Plant Science**, v. 7, p. 2024, 2017.

ZHANG, Guishan et al. Identifying circular RNA and predicting its regulatory interactions by machine learning. **Frontiers in Genetics**, v. 11, p. 655, 2020.

ZHANG, Peijing; LI, Sida; CHEN, Ming. Characterization and function of circular RNAs in plants. **Frontiers in Molecular Biosciences**, v. 7, p. 91, 2020.