

CARACTERIZAÇÃO MOLECULAR DA POPULAÇÃO VIRAL ASSOCIADA AO MOSAICO COMUM EM TRIGO NO BRASIL

Juliana Borba Valente¹, Fábio Nascimento da Silva¹, Thor Vinícius Martins Fajardo², Antônio Nhani Junior³, Douglas Lau⁴, Paulo Roberto Kuhnem Junior⁵, Fernando Sartori Pereira¹, Lucas Antônio Stempkowski¹, Monica Farias¹, Ricardo Trezzi Casa¹

¹ Universidade do Estado de Santa Catarina (UDESC), Lages, SC. E-mail: fabio.silva@udesc.br; ² Embrapa Uva e Vinho, Bento Gonçalves, RS; ³ Embrapa Informática Agropecuária, Campinas, SP; ⁴ Embrapa Trigo, Passo Fundo, RS; ⁵ Biotrigo Genética Ltda, Passo Fundo, RS.

O mosaico comum é uma importante doença viral da cultura do trigo, sendo atribuído ao *Soil-borne wheat mosaic virus* (SBWMV - *Furovirus*) e *Wheat spindle streak mosaic virus* (WSSMV - *Bymovirus*), ambos transmitidos por *Polymyxa graminis*. A doença ocasiona reduções significativas de produtividade se cultivares suscetíveis forem semeadas em áreas com a presença do vetor e em condições favoráveis como inverno frio e úmido (Lau, 2014).

Os procedimentos mais utilizados para o diagnóstico de vírus de plantas incluem testes biológicos, sorológicos e moleculares. Entretanto, experimentos com a virose do mosaico comum do trigo em áreas tritícolas brasileiras utilizando técnicas como DAS-ELISA e RT-PCR têm apresentado resultados inconclusivos, demonstrando que o conjunto de métodos para detecção dos vírus que causam mosaico necessita ser aprimorado (Carminatti et al., 2011).

A tecnologia de sequenciamento de nova geração ("*Next Generation Sequencing*", NGS), permite o desenvolvimento de ferramentas específicas e seguras para o diagnóstico de vírus (Carvalho & Silva, 2010; Roossinck et al., 2015).

Na hipótese de que o mosaico comum em trigo no Brasil seja causado por vírus divergentes dos caracterizados em populações virais de outros países, a técnica NGS é uma alternativa que pode permitir o desenvolvimento de ferramentas de diagnóstico mais consistentes. O objetivo deste trabalho foi identificar via sequenciamento de nova geração a espécie viral associada ao mosaico comum em trigo no Brasil.

Amostras de folhas e colmos de plantas de trigo com sintomas de mosaico foram coletadas em outubro de 2016 na área de caracterização de reação de

cultivares ao mosaico comum do trigo na Embrapa Trigo em Passo Fundo, Estado do Rio Grande do Sul, Brasil. Nessa área foram coletadas amostras de oito cultivares (Karl 92, Everest, OR 1, CEP 11, TBIO Toruk, TBIO Tibagi, Esporão e LG Oro), armazenadas em ultrafreezer (-80 °C). Como algumas amostras não apresentaram a quantidade necessária para extração, foram geradas amostras compostas (TBIO Toruk + TBIO Tibagi e Esporão + LG Oro). A extração de RNA de fita dupla (*double stranded RNA* - dsRNA) foi realizada conforme descrito por Valverde et al., (1990), com algumas modificações. As extrações de dsRNA foram colocadas em tubos do kit RNA stable (Biomatrica), secas em Speed Vac (Eppendorf Concentrator Plus) por 1 hora e 30 minutos, e sequenciadas por NGS (Macrogen Inc., Seul, Coréia do Sul).

As leituras (*reads*) do sequenciamento NGS foram obtidas a partir de uma biblioteca de DNA complementar (cDNA), utilizando a plataforma Illumina HiSeq 2500. A qualidade da biblioteca de cDNA foi verificada com o software FastQC e os adaptadores e leituras de baixa qualidade foram removidos com auxílio de Trimmomatic (Bolger et. al., 2014). Os dados obtidos por NGS foram analisados visando à montagem de contigs e posterior validação.

A qualidade das amostras aferida por meio de espectrometria está apresentada na Tabela 1, que mostra a relação da concentração de ácidos nucleicos em relação às proteínas contaminantes da preparação (relação A260/280 nm) e a razão A260/230 que indica a concentração de ácidos nucleicos e componentes contaminantes da preparação. A razão de absorvância A260/280 é usada para aferir a qualidade da amostra, considerada aqueda de 1,6 a 2,0 (Sambrook & Russel, 2001). Em relação a razão A260/230, valores entre 2,0-2,2 indicam boa qualidade (Wilfinger et al, 1997).

As extrações de dsRNA apresentaram uma razão A260/280 maior que 2,0 exceto a amostra 7 que apresentou 1,98. Para a relação A230/260, a amostra 1 apresentou valor acima de 2,0 (2,21), as outras amostras valores abaixo de 2,0. A amostra 5 apresentou uma relação A260/230 de 0,41 considerada muito baixa e também apresentou uma curva de absorção atípica, indicando ausência de ácidos nucleicos (Tabela 1).

As concentrações de ácidos nucleicos variaram de 641,2 a 109,2 ng/μL, as amostras 1 e 3 apresentaram os maiores valores de concentração, sendo 641,2 e 501,5 ng/μL, respectivamente (Tabela 1). Já a amostra 8 apresentou o menor valor de concentração (109,2 ng/μL). As amostras enviadas para NGS foram: (i) Karl 92; (ii)

Everest; (iii) OR 1; (iv) CEP 11; (v) TBIO Toruk + TBIO Tibagi e (vii) Esporão + LG Oro. No NGS foram geradas leituras (*reads*) as quais estão detalhadas na Tabela 2.

As amostras apresentaram um bom rendimento considerando, número total de bases sequenciadas, número total de leituras e também apresentaram conteúdo CG compatível. Adicionalmente, os parâmetros relativos a qualidade (Q20) foram superiores a 99%, os valores de “*forward*” foram superiores a 0,26%, o percentual de reverso variou de 0,01 a 0,07% e valor de descarte foi obtido apenas para a amostra Karl 92 (0,01%), indicando assim excelente qualidade do NGS.

Os dados brutos obtidos por NGS foram inicialmente analisados com o auxílio do software Trinity (Grabherr et al., 2011; Haas et al., 2013) para montagem de RNASeq. Os contigs resultantes foram analisados quanto à similaridade com sequências depositadas em bancos de dados públicos, buscando identificar sequências similares a genomas virais. Os resultados identificaram sequências homólogas a porções do genoma do trigo e outros cereais. Através do mapeamento das sequências originais no genoma do trigo (Ensembl Plants, http://plants.ensembl.org/Triticum_aestivum/Info/Index) com auxílio do software BWA (Li & Durbin, 2009), as sequências com alta similaridade a cereais foram eliminadas. As sequências não mapeadas foram novamente submetidas à montagem com o software Trinity, e paralelamente, com o software Velvet (Zerbino & Birney, 2008). A comparação dos resultados foi feita com auxílio do software Cap3 (Huang & Madan, 1999) em duas etapas: na primeira, as sequências resultantes das montagens iniciais com Trinity e Velvet foram analisadas individualmente e, na segunda, as sequências resultantes da primeira etapa foram agrupadas e novamente analisadas. Como resultado, a maior parte das sequências obtidas com Velvet estavam contidas nos *contigs* e *singlets* obtidos por meio da montagem com Trinity.

Os *contigs* e *singlets* resultantes da segunda etapa de análise com Cap3 serão subsequentemente comparados com sequências depositadas nos bancos de dados NCBI usando-se as ferramentas de pesquisa BLASTn e BLASTx com parâmetros padrão e alinhados a sequências virais. Dados preliminares apontam que alguns dos contigs têm identidade com sequências de vírus ainda não relatado na cultura do trigo. Após a obtenção do genoma viral, serão desenhados iniciadores específicos para a validação das informações obtidas no NGS em relação as amostras originais. Desta forma, será determinado o agente etiológico do mosaico comum do trigo no Brasil e

poderão ser disponibilizadas novas ferramentas de diagnóstico molecular em larga escala.

Referências

- BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v.30, p.2114-2120, 2014.
- CARMINATTI, A.J.; MAR, T.B.; LAU, D.; BIANCHIN, V.; PEREIRA, P.R.V. da S. **Avaliação de métodos para diagnose de vírus que causam Mosaico em trigo no Brasil**. VII Mostra de Iniciação Científica da Embrapa Trigo Resumos. Documentos online 133. Disponível em <www.embrapa.br/trigo>. Acesso: 20 de abril de 2018.
- CARVALHO, M.C.C.G.; SILVA, D.C.G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, v.40, p.735-744, 2010.
- GRABHERR, M.G.; HAAS, B.J.; YASSOUR, M.; LEVIN, J.Z.; THOMPSON, D.A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q.; CHEN, Z.; MAUCELI, E.; HACOEN, N.; GNIRKE, A.; RHIND N.; PALMA, F.; BIRREN, B.W.; NUSBAUM, C.; LINDBLAD-TOH, K.; FRIEDMAN, N.; REGEV, A. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **Nature Biotechnology**, v.29, p.644-652, 2011.
- HAAS, B.J.; PAANICOLAOU, A.; YASSOUR, M.; GRABHERR, M.; BLOOD, P.D.; BOWDEN, J.; COUGER, M.B; ECCLES, D.; LI, B.; LIEBER, M.; MACMANES, M.D.; OTT, M.; ORVIS, J.; POCHE, N.; STROZZI, F.; WEEKS, N.; WESTERMAN, R.; WILLIAM, T.; DEWEY, C.N.; HENSCHEL, R.; LEDUC, R.D.; FRIEDMAN, N.; REGEV, A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature Protocols**, v.8, p.1494-512, 2013.
- HUANG, X; MADAN, A. CAP3: A DNA Sequence Assembly Program. **Genome Research**, v.9, p.868-887, 1999.
- LAU, D. Como manejar as principais viroses que atacam a cultura do trigo. **Cultivar Grandes Culturas**, v.185 p.32-36, 2014.
- LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. **Bioinformatics**, v.25, p.1754-1760, 2009.
- ROOSSINCK, J.M.; MARTIN, P.D.; ROUMAGNAC, P. Plant Virus Metagenomics: Advances in Virus Discovery. **Phytopathology**, v.105, p.716-727, 2015.

SAMBROOK, J.E.; RUSSELL, D.W. **Molecular cloning: a laboratory manual**. 3 ed. New York: Cold Spring Harbor Laboratory Press, 2100pp, 2001.

VALVERDE, R.A.; NAMETH, S.T.; JORDAN, R.L. Analysis of double-stranded RNA for plant virus diagnosis. **Plant Disease**, v.74, p.255-258, 1990.

WILFINGER, W.W.; MACKEY, K.; CHOMCZYNSKI, P. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity.

BioTechniques, v.22, p.478-481, 1997.

ZERBINO, D.R.; BIRNEY, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, v.18, p.821-829, 2008.

Tabela 1. Avaliação qualitativa e quantitativa das extrações de dsRNA.

| Amostra | Concentração (ng/μL) | A260/A280* | A260/A230** |
|---------|----------------------|------------|-------------|
| 1 | 641,2 | 2,16 | 2,21 |
| 2 | 378,1 | 2,22 | 1,12 |
| 3 | 501,5 | 2,19 | 1,60 |
| 4 | 175,9 | 2,07 | 1,58 |
| 5 | 139,0 | 2,04 | 0,41 |
| 6 | 149,2 | 2,12 | 1,50 |
| 7 | 169,4 | 1,98 | 1,26 |
| 8 | 109,2 | 2,09 | 1,49 |

*Razão de absorvância do ácido nucleico das amostras extraídas e proteínas contaminantes.

**Razão de absorvância do ácido nucleico e componentes contaminantes.

Tabela 2. Dados brutos obtidos a partir do sequenciamento de nova geração.

| Amostra | Total de leituras de bases (pb)* | Total de leituras** | GC %*** | AT %**** |
|---------|----------------------------------|---------------------|---------|----------|
| 1 | 3.079.122.966 | 30.486.366 | 51,73 | 48,27 |
| 2 | 2.894.869.272 | 28.662.072 | 51,11 | 48,88 |
| 3 | 2.890.716.556 | 28.620.956 | 51,38 | 48,61 |
| 4 | 3.361.638.954 | 33.283.554 | 51,74 | 48,25 |
| 6 | 2.956.241.518 | 29.269.718 | 49,48 | 50,52 |
| 7 | 3.020.315.716 | 29.904.116 | 49,70 | 50,29 |

*Total de leitura de bases : Número total de bases sequenciadas. **Total de leituras: Número total de leituras (reads). ***GC (%): conteúdo GC. ****AT (%): conteúdo AT.