CABI Agriculture and Bioscience

SHORT COMMUNICATION

Open Access

Effects of tissue type and season on the detection of regulated sugarcane viruses by high throughput sequencing

Stephen Bolus[1], Kate Wathen-Dunn[2], Samuel C. Grinstead[1], Xiaojun Hu[3], Martha Malapi[4] and Dimitre Mollov[5*]

Abstract

High throughput sequencing (HTS) can supplement and may replace diagnostic tests for plant pathogens. However, the methodology and processing of HTS data must first be optimized and standardized to ensure the sensitivity and repeatability of the results. Importation of sugarcane into the United States is highly regulated, and sugarcane plants are subjected to strict quarantine measures and diagnostic testing, especially for the presence of certain viruses of regulatory concern. Here, we tested whether HTS could reliably detect four RNA and three DNA sugarcane viruses over three seasons (fall, winter, and spring) and in three tissue types (root, stem, and leaves). Using HTS on ribosomal depleted total RNA samples, we reliably detected RNA viruses in all tissue types and across all seasons, but we failed to confidently detect DNA viruses in some samples. We recommend that future optimization be employed to ensure the robust and reliable detection of all regulated sugarcane viruses by HTS.

Keywords Sugarcane, Virus, High throughput sequencing, RNA-Seq, Season, Tissue

Background

Sugarcane (*Saccharum* spp.) is grown in subtropical and tropical regions worldwide and is an important crop for food and energy use. Sugarcane is clonally propagated, and the global exchange of germplasm carries risk for the subsequent spread of plant viruses. Many viruses are known to infect sugarcane plants, often reducing their yield (Putra et al. 2014; ElSayed et al. 2015). In the United States, several methods are utilized to detect known viruses of regulatory concern in imported sugarcane materials. High throughput sequencing (HTS) is a method for plant pathogen detection that is gaining support as it becomes less expensive (Maree et al. 2018; Villamor et al. 2019). This technology can supplement existing detection methods by identifying previously unknown or highly divergent pathogen species and could eventually replace existing detection methods and improve quarantine measures (Maree et al. 2018; Villamor et al. 2019). However, we must first explore the limits of this technology in detecting viruses.

Main text

Sugarcane viruses of regulatory concern in the United States include DNA and RNA viruses. In this study, we focused on the detection of the following RNA viruses: sugarcane yellow leaf virus (SCYLV, genus *Polerovirus*), Fiji disease virus (FDV, genus *Fijivirus*), sugarcane striate

*Correspondence:
Dimitre Mollov
dimitre.mollov@usda.gov
[1] USDA-ARS, National Germplasm Resources Laboratory, Beltsville, MD, USA
[2] Sugar Research Australia Limited, Indooroopilly, QLD, Australia
[3] USDA-APHIS Plant Germplasm Quarantine Program, Beltsville, MD, USA
[4] USDA-APHIS-BRS, Biotechnology Risk Analysis Programs, Riverdale, MD, USA
[5] USDA-ARS, Horticultural Crops Disease and Pest Management Research Unit, Corvallis, OR, USA

mosaic-associated virus (SCSMaV, genus *Sustrivirus*), and sugarcane streak mosaic virus (SCSMV, genus *Poacevirus*) as well as the following DNA viruses: sugarcane bacilliform virus (SCBV, unclassified *Badnavirus*), sugarcane white streak virus (SCWSV, genus *Mastrevirus*), and sugarcane streak Egypt virus (SCSEV, genus *Mastrevirus*). We included plants infected with only one virus and other plants known to be infected with two and three viruses (Fig. 1, Additional file 2: Table S1), as co-infection status could affect virus detection (Syller 2012).

How tissue type affects virus detection in sugarcane is largely unexplored. FDV and SCYLV were both previously detected in root and leaf tissues by ELISA and tissue blots, respectively (Wagih and Adkins 1996; Lehrer et al. 2007). In the United States, virus testing for regulatory purposes is currently performed on leaf samples

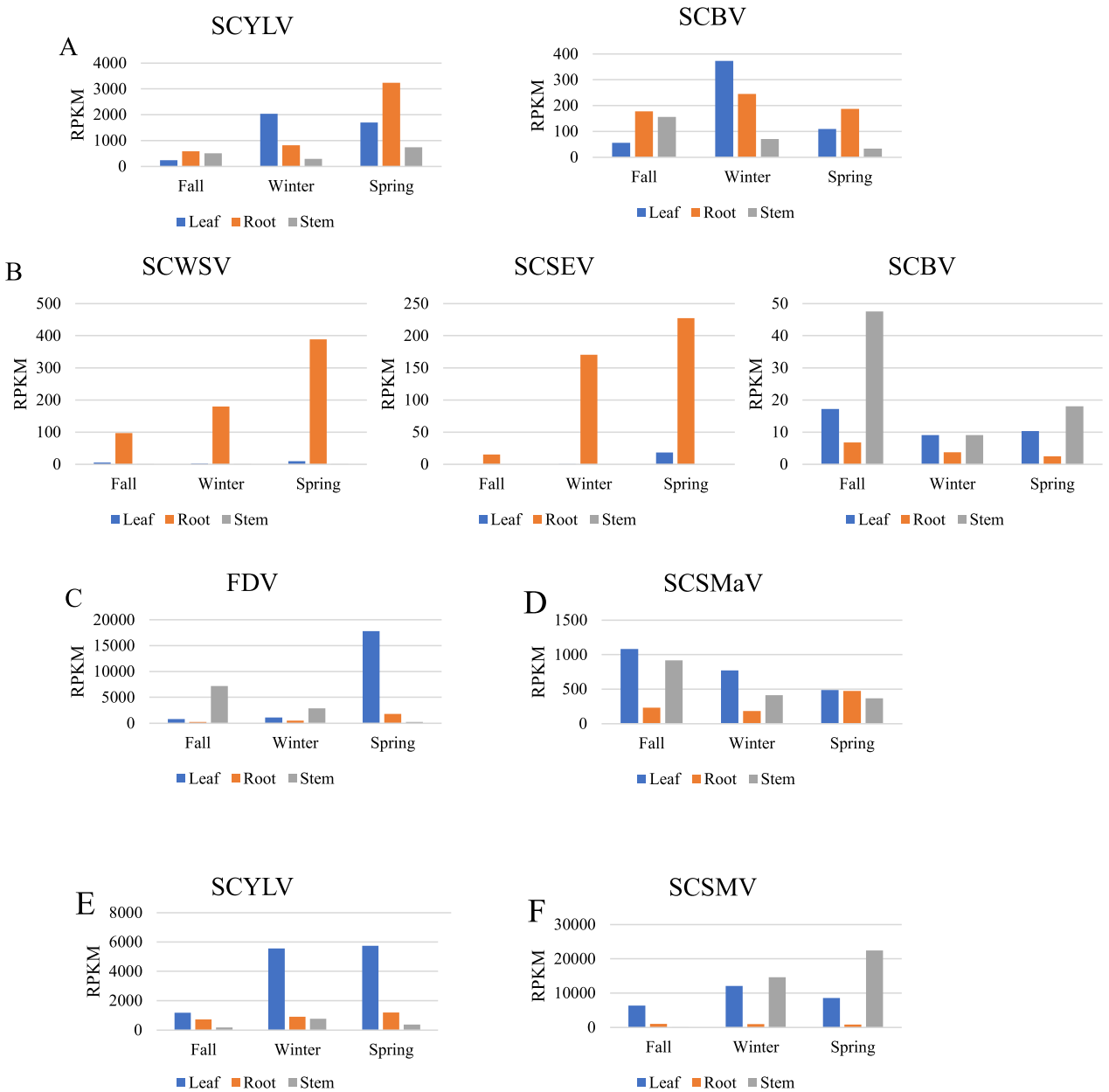

Fig. 1 Bar graphs showing HTS results in reads per kilobase per million reads (RPKM) measurements for seven sugarcane viruses in six plants across tissue types and seasons. **A** plant P1, **B** plant P2, **C** plant P3, **D** plant P4, **E** plant P5, and **F** plant P6. SCYLV: sugarcane yellow leaf virus; SCBV: sugarcane bacilliform virus; SCWSV: sugarcane white streak virus; SCSEV: sugarcane streak Egypt virus; FDV: Fiji disease virus; SCSMaV: sugarcane striate mosaic-associated virus; SCSMV: sugarcane streak mosaic virus

taken from sugarcane plants grown under quarantine greenhouse conditions during fall and spring seasons. The results from a related study on the effect of fall and spring seasons on the detection of sugarcane viruses of regulatory concern suggested that spring may be the optimum season for virus detection by HTS using leaf samples (Malapi-Wight et al. 2021). We sought to expand on these findings and look at whether virus detection by HTS fluctuated across three seasons (spring, fall and winter) and three plant tissues (leaves, roots and stems).

We collected leaves, roots, and stems from six greenhouse-grown (16/8 h day/night) sugarcane plants in September 2019 (fall), December 2019 (winter), and April 2020 (spring) in Beltsville, Maryland, USA. These plants are diverse in genetic background, are infected with viruses of regulatory concern, and are routinely used as positive controls for testing in the USDA-APHIS Sugarcane Quarantine Program (Malapi-Wight et al. 2021). From the collected samples, we extracted RNA using RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions, except the lysate was further processed by incubating at 70 °C for 10 min before loading on the QIAshredder spin column. RNA samples were purified using the Monarch RNA Cleanup kit (NEB, MA, USA) as necessary. We outsourced DNA and rRNA depletion, cDNA library preparation, and sequencing on an Illumina NextSeq 500 platform as single end 75 bases reads (SeqMatic, CA, USA). The raw reads were trimmed using Trimmomatic (v0.36) (Bolger et al. 2014), and the remaining rRNA reads were subtracted using bbduk.sh in BBMap (v38.90) (Bushnell 2014). These cleaned reads were then assembled using SPAdes (v3.13.0) with default parameters (Bankevich et al. 2012). Contigs from SPAdes were annotated using Blastn (v2.10.1+) (Camacho et al. 2009) against NCBI viral reference database (Brister et al. 2015) and DIAMOND (v2.0.9) Blastx (Buchfink et al. 2015) against Reference Viral Databases RVDB (v18.0) (Bigot et al. 2019). The closest viral sequence with E-value = 0.0 was identified as a reference virus, which was used to map reads using BWA (v0.7.17-r1188) with parameters: -k 12 -A 1 -B 3 -O 1 -E 1 (Li and Durbin 2009). Reads per kilobase per million reads (RPKM) for each virus were calculated by multiplying the number of mapped reads to a reference virus isolate sequence (Additional file 2: Table S1) by 10^9, and subsequently dividing this number by the product of the total number of trimmed and rRNA subtracted reads from that sample and the nucleotide length of the reference virus isolate sequence (Wagner et al. 2012). For presentation in Additional file 3: Table S2, the RPKM measurements were then rounded to the nearest whole number. For plant 3 infected with FDV, individual RPKM values were calculated for each of the 10 virus segments, and the average RPKM over the 10 virus segments was used for data analysis and presentation purposes.

To look at general patterns in the detection of regulated sugarcane viruses based on the tissue and season of sample collection, RPKM values were first averaged over each individual plant as some plants were co-infected with two or three viruses. The averages and standard errors for the RPKM averages from all six plants were plotted on a graph (Additional file 1: Fig. S1). Overall, there was high variability in the detection of regulated viruses in these different sugarcane plants over season and tissue types. Although there was a trend for higher virus detection in spring and leaf tissues, observations in the raw data (Fig. 1, Additional file 3: Table S2) suggest that this trend was largely influenced by FDV in plant P3. FDV was previously identified at much higher levels by HTS in spring versus fall leaf tissue (Malapi-Wight et al. 2021). Overall, there was no major effect of season and tissue type on the detection of all sugarcane viruses in our study (statistical analyses not presented).

From a practical standpoint, we were interested in how the RPKM values we observed in our data would translate to the depth of sequencing required to confidently diagnose these samples as positive for virus(es) of regulatory concern. To accomplish this aim, we used the fastq-tools package (version 0.8.3) to make 3 replications of random, sub-sampled sets of 0.5, 1, 5, 10, 20, and 25 million reads from our sequencing data. We chose these discrete sets to be consistent with those explored in Malapi-Wight et al. 2021. The replications were mapped and summarized using the BBMap package (version 38.73). We assigned thresholds of ≥ 60% reference genome coverage for DNA viruses and ≥ 80% reference genome coverage for RNA viruses, and we listed the lowest million reads sub-sampled set where all three replications fulfilled these threshold criteria for each virus/plant sample in Additional file 3: Table S2. Using our methods for detection by HTS, all RNA sugarcane viruses of regulatory concern studied here were confidently detected at or less than 5 million reads in all samples and across all tissue types (Additional file 3: Table S2). Our findings are consistent with those of a previous study on ribosomal RNA depleted total RNA samples from spring and fall sugarcane leaves, where there was confident detection of various RNA viruses in all samples by HTS (Malapi-Wight et al. 2021). The DNA virus SCBV was confidently detected at one million reads or less in plant P1, a plant co-infected with SCBV and SCYLV, but it was not detected or detected between five and 20 million reads in samples from different tissue types and seasons in plant P2, a plant co-infected with SCWSV, SCSEV, and SCBV (Additional file 3: Table S2). Since we did not control for host genotype nor virus isolate in this study, it is hard to

discern whether these differences in detection are due to co-infection status, virus isolate differences, or differences in the host plant genotype. We also failed to confidently detect SCWSV and SCSEV in plant P2, except for in root tissues and some spring leaf tissues (Additional file 3: Table S2). Sugarcane DNA viruses were previously reported to be more difficult to detect than RNA viruses by HTS using ribosomal RNA depleted total RNA from spring and fall leaves (Malapi-Wight et al. 2021). Interestingly, the highest RPKM values were observed in root tissues for DNA viruses SCWSV and SCSEV (Additional file 3: Table S2, Fig. 1). Although these observations are from a single, co-infected plant, our HTS data (based on RNA-Seq) suggest that the DNA viruses SCWSV and SCSEV may be expressed more in root tissues (Additional file 3: Table S2, Fig. 1). To further optimize the use of HTS for the detection of all sugarcane viruses of regulatory concern, we suggest performing HTS on other nucleotide purifications, such as DNA or small RNA, in addition to ribosomal depleted total RNA. Small RNA sequencing was shown to outperform ribosomal RNA depleted total RNA sequencing for the detection of some single stranded DNA viruses and viroids (Pecman et al. 2017).

Lastly, since our sequencing data included transcriptomic data from sugarcane samples, we were interested in analyzing the differences in sugarcane gene expression across our samples. Files were trimmed using Trimmomatic (v0.36) (Bolger et al. 2014), and the quality was analyzed using FastQC (FastQC v0.11.5). Trimmed files were imported into CLC Genomics Workbench v12.0 (CLCGxWb) (Qiagen) and were mapped to the PacBio Iso-Seq sugarcane transcriptome (Hoang et al. 2017) using the 'RNAseq Analysis' function in CLCGxWb with the following parameters: batch mapping, local alignment, mismatch = 2, insertion = 3, deletion = 3, length fraction = 0.8, similarity fraction = 0.8, map both strands, and maximum number of hits for a read = 20. Differentially expressed genes were identified using CLCGxWb. The effects of each variable (treatment, tissue, genotype, and season) on gene expression were each analyzed separately across groups (ANOVA-like). The differentially expressed genes were further filtered for those with Bonferroni ≤ 0. An online resource (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to construct a Venn diagram (Fig. 2) of the differentially expressed genes identified in each analysis to see how they interacted/overlapped.

Most of the differentially expressed genes (3,643) identified were shared between plant genotypes and treatments (virus infection status; Fig. 2). Thirty additional genes were differentially expressed between genotypes, treatments, and tissues, and two genes were differently

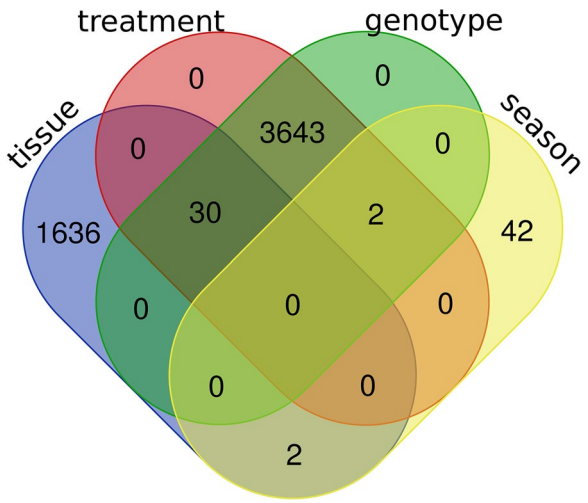

Fig. 2 Venn diagram showing the number of differentially expressed sugarcane genes by sample treatment (viral pathogen infection), season, tissue, and genotype

expressed between genotypes, treatments, and seasons. Since the plants used in our experiment were different genotypes of sugarcane and were infected with different viruses, and in some plants, combinations of viruses, it was not possible to parse out differences in gene expression attributed to genotype or virus infection status (treatment). Future experimentation controlling for either plant genotype or virus infection status could help identify specific plant responses to viral infection. Across the different plants, more differently expressed genes were identified between tissues (1,636) than between seasons (42), and two genes were identified as differently expressed between both seasons and tissues. These findings are largely expected, given the vast differences in gene expression expected between plant tissue types, and the fact that our sugarcane plants were grown under greenhouse conditions. Nevertheless, these data support the conclusion that our sugarcane plants uniformly had more gene expression differences based on tissue type than seasonality, indicating that under quarantine greenhouse conditions, future optimization work should likely focus more on tissue type detection differences rather than seasonal differences.

Conclusions

In conclusion, we confidently detected all selected sugarcane RNA viruses of regulatory concern across seasons and tissue types. However, by performing HTS on ribosomal depleted total RNA samples, we failed to detect certain sugarcane DNA viruses of regulatory concern in some samples. We hope that this preliminary work sheds light on potential limitations of HTS pipelines and will

help inform future optimization studies of HTS protocols for the robust and reliable detection of all sugarcane viruses of regulatory concern.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43170-023-00175-1.

Additional file 1: Figure S1. General trends for the detection of sugarcane viruses of regulatory concern using high throughput sequencing. RPKM = reads per kilobase per million reads; F = Fall; S = Spring; W = Winter; l = leaf; st = stem; and r = root. Error bars represent standard error based on six sugarcane plants.

Additional file 2: Table S1. Accession numbers from the National Center for Biotechnology Information (NCBI) used for read mapping in reads per kilobase per million reads (RPKM) measurements and depth of sequencing analysis.

Additional file 3: Table S2. Detection of regulated sugarcane viruses as measured by reads per kilobase per million reads (RPKM) values and depth of sequencing analysis.

Acknowledgements
Not applicable.

Author contributions
All authors have read and agreed to the published version of the manuscript. Conceptualization: DM, MM; Methodology: KWD, SCG, XH, SB, DM; Formal analysis and investigation: SB, KWD, SCG, XH; Writing, original draft preparation: SB, DM; Writing, review and editing: SB, KWD, SCG, XH, MM, DM; Funding acquisition: DM; Resources: DM, MM; Supervision: DM.

Funding
This work was supported by U.S. Department of Agriculture, Agricultural Research Service, Research Project 8042-22000-302-00-D.

Availability of data and materials
The datasets generated and/or analyzed during the current study are not publicly available due to the sensitivity of the quarantine materials used in this study but are available from the corresponding author on reasonable request.

Declarations

Competing interests
Kate Wathen-Dunn was employed by Sugar Research Australia Limited. The remaining authors have no conflicts of interest to declare that are relevant to the content of this article.

Received: 21 June 2023 Accepted: 29 August 2023
Published online: 13 September 2023

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77. https://doi.org/10.1089/cmb.2012.0021.

Bigot T, Temmam S, Pérot P, Eloit M. RVDB-prot, a reference viral protein database and its HMM profiles. F1000Research. 2019. https://doi.org/10.12688/f1000research.18776.2.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. Nucleic Acids Res. 2015;43(D1):D571–7. https://doi.org/10.1093/nar/gku1207.

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60. https://doi.org/10.1038/nmeth.3176.

Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Berkeley: Lawrence Berkeley National Lab (LBNL); 2014.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinform. 2009;10:1–9. https://doi.org/10.1186/1471-2105-10-421.

ElSayed AI, Komor E, Boulila M, Viswanathan R, Odero DC. Biology and management of sugarcane yellow leaf virus: an historical overview. Adv Virol. 2015;160:2921–34. https://doi.org/10.1007/s00705-015-2618-5.

Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJ. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. BMC Genomics. 2017;18(1):1–22. https://doi.org/10.1186/s12864-017-3757-8.

Lehrer AT, Schenck S, Yan SL, Komor E. Movement of aphid-transmitted Sugarcane yellow leaf virus (ScYLV) within and between sugarcane plants. Plant Pathol. 2007;56(4):711–7. https://doi.org/10.1111/j.1365-3059.2007.01557.x.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Malapi-Wight M, Adhikari B, Zhou J, Hendrickson L, Maroon-Lango CJ, McFarland C, Foster JA, Hurtado-Gonzales OP. HTS-based diagnostics of sugarcane viruses: Seasonal variation and its implications for accurate detection. Viruses. 2021;13(8):1627. https://doi.org/10.3390/v13081627.

Maree HJ, Fox A, Al Rwahnih M, Boonham N, Candresse T. Application of HTS for routine plant virus diagnostics: State of the art and challenges. Front Plant Sci. 2018;9:1082. https://doi.org/10.3389/fpls.2018.01082.

Pecman A, Kutnjak D, Gutiérrez-Aguirre I, Adams I, Fox A, Boonham N, Ravnikar M. Next generation sequencing for detection and discovery of plant viruses and viroids: Comparison of two approaches. Front Microbiol. 2017;8:1998. https://doi.org/10.3389/fmicb.2017.01998.

Putra LK, Kristini A, Achadian EM, Damayanti TA. Sugarcane streak mosaic virus in Indonesia: distribution, characterisation, yield losses and management approaches. Sugar Tech. 2014;16:392–9. https://doi.org/10.1007/s12355-013-0279-9.

Syller J. Facilitative and antagonistic interactions between plant viruses in mixed infections. Mol Plant Pathol. 2012;13(2):204–16. https://doi.org/10.1111/j.1364-3703.2011.00734.x.

Villamor DE, Ho T, Al Rwahnih M, Martin RR, Tzanetakis IE. High throughput sequencing for plant virus detection and discovery. Phytopathology. 2019;109(5):716–25. https://doi.org/10.1094/phyto-07-18-0257-rvw.

Wagih MW, Adkins SW. Detection of Fiji disease virus in sugarcane by ELISA. Aust J Exp Agric. 1996;36(1):105–10. https://doi.org/10.1071/EA9960105.

Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131:281–5. https://doi.org/10.1007/s12064-012-0162-3.

Publisher's Note