

RESEARCH NOTE

Open Access



# High-throughput sequencing of virus-infected *Cucurbita pepo* samples revealed the presence of Zucchini shoestring virus in Zimbabwe

Charles Karavina<sup>1</sup>, Jacques Davy Ibaba<sup>2\*</sup>  and Augustine Gubba<sup>2</sup>

## Abstract

**Objectives:** Plant-infecting viruses remain a serious challenge towards achieving food security worldwide. Cucurbit virus surveys were conducted in Zimbabwe during the 2014 and 2015 growing seasons. Leaf samples displaying virus-like symptoms were collected and stored until analysis. Three baby marrow samples were subjected to next-generation sequencing and the data generated were analysed using genomics technologies. Zucchini shoestring virus (ZSSV), a cucurbit-infecting potyvirus previously described in South Africa was one of the viruses identified. The genomes of the three ZSSV isolates are described analysed in this note.

**Results:** The three ZSSV isolates had the same genome size of 10,297 bp excluding the polyA tail with a 43% GC content. The large open reading frame was found at positions 69 to 10,106 on the genome and encodes a 3345 amino acids long polyprotein which had the same cleavage site sequences as those described on the South African isolate except for the P1-pro site. Genome sequence comparisons of all the ZSSV isolates showed that the isolates F7-Art and S6-Prime had identical sequence across the entire genome while sharing 99.06% and 99.34% polyprotein nucleotide and amino acid sequence identities, respectively with the isolate S7-Prime.

**Keywords:** Next generation sequencing, Potyvirus, Plant virus, Cucurbit, Zimbabwe, Zucchini shoestring virus, High-throughput sequencing

## Introduction

Cucurbit is a generic term used to denote all species within the Family *Cucurbitaceae* also known as the gourd family [1]. Numerous cucurbit crops are economically important worldwide. Cucurbits are consumed in different ways as fruits or vegetables, providing essential nutrients and dietary fibre [2]. In Zimbabwe, some of the cultivated cucurbits include the cucumber (*Cucumis melo* L.), the watermelon (*Citrullus lanatus* (Thunb.) Matsum.

& Nakai), the melon (*Cucumis melo* L.), the pumpkin (*Cucurbita maxima* Duch.), the butternut (*Cucurbita moschata* Duch.) and the baby marrow (*Cucurbita pepo* L.). They are widely grown by both commercial and smallholder farmers as food and cash crops. Virus diseases on cucurbits produce diverse symptoms that result in yield reduction and in severe instances compromised fruit quality [3, 4]. The negative effects of plant-infecting viruses on crops are more prominent especially in countries where their studies are underdeveloped.

High-throughput sequencing (HTS), also called next-generation sequencing (NGS) describes a series of technologies whereby millions or billions of DNA molecules are sequenced simultaneously [5]. The application of

\*Correspondence: janick.ibaba@gmail.com

<sup>2</sup> Discipline of Plant Pathology, University of KwaZulu-Natal, Agriculture Campus, 1 Agric Avenue, Scottsville, Pietermaritzburg 3209, Republic of South Africa

Full list of author information is available at the end of the article



these ever-growing sequencing technologies and bioinformatics data analysis to the studies of plant-infecting viruses, which started in 2009 [5], have revolutionized the fields of virus discovery and diagnostics, resulting in unprecedented virus discoveries from any host and environment [6]. Unlike other popular techniques such as the enzyme-linked immunosorbent assay, molecular hybridization and polymerase chain reaction that mainly work on known pathogens, HTS data analysis has made possible the identification of sequences of known or unknown viruses from any host without any prior knowledge of the disease aetiology [7, 8].

Zucchini shoestring virus (ZSSV) was discovered among other known cucurbit-infecting viruses in 2015 in South Africa when the RNA from severely distorted Baby marrow leaves were subjected to HTS [9, 10]. Genomics and taxonomic studies revealed that ZSSV is a new species in the genus *Potyvirus* [10]. The International Committee TV subsequently ratified these findings [11]. The genus *Potyvirus* is one of the 8 genera that composed the family *Potyviridae*. Members in that family, also known as potyvirids, are differentiated by the host range, genomic features and phylogeny, with a species demarcation criterion set to a nucleotide and amino sequence identity less than 76% and 82%, respectively for the large open reading frame (ORF) or its protein product. In instances where the complete ORF sequence is not available, similar criteria can be used for the coat protein (CP) coding region [12].

Viruses that belong to the genus *Potyvirus* have non-enveloped, flexuous and filamentous virions of 680–900 nm in length and 11–20 nm in diameter. The genome of potyviruses is a positive-sense ssRNA molecule with its 5' terminus covalently linked to the viral protein genome linked (VPg) and its 3' end polyadenylated. The 10,000 bp genome harbours two ORFs that encode eleven multifunctional proteins. A large ORF is translated into a single polyprotein that is cleaved at semi-conserved sites by three self-encoded proteases into ten mature proteins namely the protein 1 protease (P1-Pro), the helper component proteinase (HC-Pro), Protein 3 (P3), six kilodalton peptide 1 (6K1), the 6K2, the cytoplasmic inclusion (CI), the nuclear inclusion A protease (NIa-Pro), the nuclear inclusion B RNA-dependent RNA polymerase (NIb), the VPg and the CP [12]. A smaller ORF, named the pretty interesting *Potyviridae* ORF (PIPO), is generated by a polymerase slippage mechanism and is expressed as the trans-frame protein P3N-PIPO [13–15].

In this note, we described and studied the genome sequences of three ZSSV isolates obtained through HTS of infected baby marrow leaves collected in Zimbabwe.

## Main text

### Sample sources

Virus surveys were conducted in selected cucurbit farms in Harare, Zimbabwe, in 2014 and 2015 growing seasons. Baby marrow plants (*Cucurbita pepo*) displaying mosaic and mild leaf distortion (Fig. 1) were the most prevalent symptoms of viral aetiology observed throughout the surveys. Labelled samples were collected and consisted of one symptomatic younger leaf fully developed preserved in RNAlater Solution (ThermoFisher Scientific, USA). Three leaf samples from three different farms were randomly selected for HTS.

### High-throughput sequencing and data analysis

Total RNA was extracted from each leaf sample using the Quick-RNA Miniprep Kit (Zymo Research, USA) as per the manufacturer's instructions and was shipped on dry ice to the Agricultural Research Council Biotechnology Platform (ARC-BTP) in Pretoria, South Africa for sequencing on the HiSeq platform (Illumina Inc., USA). For each sample, the data generated from sequencing was analysed as follows. The read quality was assessed using



**Fig. 1** Picture of the most common symptom observed on baby marrow plants during the survey conducted in selected cucurbit-growing farms in Harare in 2014 and 2015

FastQC version 0.11.5 (Babraham Bioinformatics) and when necessary, Trimmomatic version 0.36 [16] was used to trim. De novo assembly was then performed using SPAdes version 3.10.1 [17] according to the developer’s instructions. Nucleotide blast was performed on all contig using BLAST+ [18].

**Genomics and phylogenetic analysis**

The ORFfinder web version (<https://www.ncbi.nlm.nih.gov/orffinder/>) was used to identify ORFs. ClustalW [19] was used to do multiple sequence alignment. Nucleotide and amino acid sequence identities were performed online with SIAS (<http://imed.med.ucm.es/Tools/sias.html>). MEGA X software version 10.1.7 [20] was used to find the best evolutionary model fitting our phylogenetic analysis and to infer the maximum likelihood tree accordingly. ZSSV being one of the species in the “Papaya ringspot virus (PRSV) cluster” of cucurbit-infecting potyviruses, the phylogenetic analyses were performed using the CP coding sequences of selected members of this cluster.

**Results**

**ZSSV genome sequence identified from HTS data analysis**

The BLAST results identified one contig from each sample as a perfect match to the full-length genome sequence of the South African (SA) ZSSV isolate (GenBank accession number: KU355553.1). These sequences were then referred as ZSSV isolates F7-Art, S6-Prime and S7-Prime. The coverage values were 30x, 66x and 80x for F7-Art, S6-Prime and S7-Prime, respectively. The genome size was the same for the three isolates and consisted of 10,297 bp excluding the polyA tail with GC contents varying between 42.92 and 42.96%. Each isolate sequence was submitted to GenBank and was given accession number as surmised in Table 1.

**ZSSV genome analysis and phylogeny**

The genome features common to the three isolates included the lengths and the positions of both ORFs and the polyprotein cleavage site sequences. The large ORF was located at positions 69 to 10,106 of the genome.

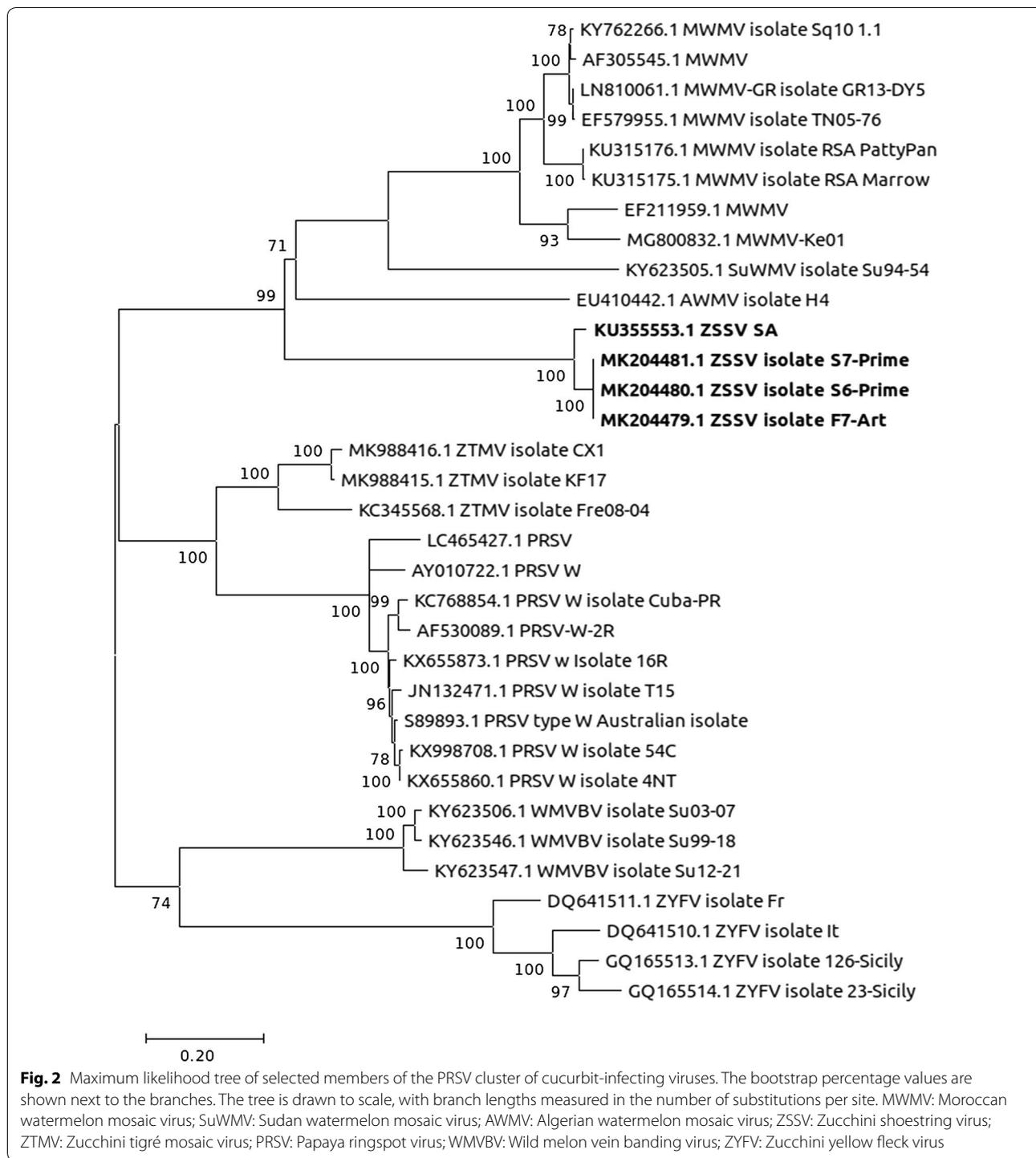
The polyprotein resulting from the direct translation of the large ORF was 3345 amino acids long. The PIPO ORF was situated from nucleotide position 3611 to 3793. The LAIGN box that has been reported to play a role in virus movement and amplification [21] and the FRNK box involved in RNA silencing and symptom development [22] were identified on the HC-Pro of all the ZSSV isolates. The motifs DAG [23], RITC and PTR involved in aphid transmission were also part of the CP and the HC-Pro.

The polyprotein cleavage site sequences of the three isolates described in this study were the same as the SA isolate [10] except for the P1-pro site that was IVHY|S instead of IIHY|S. Genome sequence comparisons of all the ZSSV isolates are available in Additional files 1 and 2. They showed that the isolates F7-Art and S6-Prime had identical sequence across the entire genome while sharing 99.06% and 99.34% polyprotein nucleotide and amino acid sequence identities, respectively with the isolate S7-Prime. The CP, 6K1, 6K2 and 5’ terminus nucleotide and amino acid sequences were the same for the three isolates under study. The amino acid sequence of the HC-Pro and the NIa-Pro were 100% identical although their corresponding nucleotide sequences were not. The lowest percentage values of 97.78% and 97.21% were recorded with the P1-Pro nucleotide and amino acid sequence, respectively. When compared with the SA isolate, the polyprotein nucleotide sequence identities was 91.08% with the isolates F7-Art and 92.02 with the isolate S7-Prime. The polyprotein amino acid sequence identities percentages were a bit higher at 95.84% and 96.5% against the isolates F7-Art and the isolate S7-Prime, respectively. At the individual genome features nucleotide and amino acid sequence identity between the SA isolate and the ZSSV isolates from Zimbabwe ranged from 87.87 to 96.39% and from 87.1 to 99.34%, respectively.

The phylogenetic analysis involved 33 nucleotide sequences and was inferred using the general time-reversible model with a discrete Gamma distribution (5 categories (+G, parameter = 0.8565)) and invariable sites ([+I], 27.21% sites). The tree with superior log-likelihood value (-9554.87) was automatically selected (Fig. 2). The selected isolates in the tree were divided into three main groups. One group was made of Moroccan watermelon mosaic virus (MWMV) isolates, Sudan watermelon mosaic virus (SuWMV) isolates, Algerian watermelon mosaic virus (AWMV) isolates and ZSSV isolates. In another group were included Zucchini tigré mosaic virus (ZTMV) isolates and PRSV isolates. The last group comprised Wild melon vein banding virus (WMVBV) isolates and Zucchini yellow fleck virus (ZYFV) isolates. All the ZSSV isolates clustered together with 100% bootstrap value.

**Table 1 GenBank accession number of the ZSSV isolates described in this study**

Isolate name	Genbank accession number
Zucchini shoestring virus isolate F7-Art	MK204479.1
Zucchini shoestring virus isolate S6-Prime	MK204480.1
Zucchini shoestring virus isolate S7-Prime	MK204481.1



**Discussion**

PRSV cluster of cucurbit-infecting virus include eight acknowledged species. Four of those species, ZTMV [24], ZSSV [10], SuWMV and WMVBV [25], have been reported in the past 7 years. Moreover, MWMV, AWMV, ZSSV, SuWMV and WMVBV were identified in Africa,

suggesting that the PRSV cluster underwent an important diversification in Africa [25]. Out of these viruses present in Africa, MWMV is the widespread one having been reported in all African regions [3, 26–32]. The HTS in this study made the detection of ZSSV on infected leaf sample possible. The presence of ZSSV in cultivated baby

marrow plants from the surveyed farms may indicate either a broader geographical distribution of the virus or its spreading across borders. The occurrence of ZZSV in Zimbabwe highlights the need to conduct further studies on its epidemiology and to develop effective management strategies.

## Limitations

1. The small number of samples analysed in that study was one of the limitations.
2. ZSSV at this stage of the study can not be considered the main causal agent of the symptoms identified in the virus surveys.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13104-020-4927-3>.

**Additional file 1.** Nucleotide sequence identities of the Zucchini shoestring virus (ZSSV) isolates. Table displaying the nucleotide sequence identities in percentage between all ZSSV isolates available on GenBank.

**Additional file 2.** Amino acid sequence identities of the Zucchini shoestring virus (ZSSV) isolates. Table displaying the amino acid sequence identities in percentage between all ZSSV isolates available on GenBank.

## Abbreviations

AWMV: Algerian watermelon mosaic virus; ARC-BTP: Agricultural Research Council Biotechnology Platform; CI: Cytoplasmic inclusion; CP: Coat protein; HC-Pro: Helper component proteinase; HTS: High-throughput sequencing; MWMV: Moroccan watermelon mosaic virus; NGS: Next generation sequencing; Nla-Pro: Nuclear inclusion A protease; Nlb: Nuclear inclusion B RNA-dependent RNA polymerase; ORF: Open reading frame; P1-Pro: Protein 1 protease; P3: Protein 3; PIPO: Pretty interesting *Potyviriidae* open reading frame; PRSV: Papaya ringspot virus; RNA: Ribonucleic acid; SA: South Africa; SuWMMV: Sudan watermelon mosaic virus; VPg: Viral protein genome-linked; WMVBV: Wild melon vein banding virus; ZTMV: Zucchini tigré mosaic virus; ZSSV: Zucchini shoestring virus; ZYFV: Zucchini yellow fleck virus; 6k1: Six kilodalton peptide 1; 6k2: Six kilodalton peptide 2.

## Acknowledgements

The authors will like to acknowledge the ARC-BTP technical team for providing good HTS services. Charles Karavina expresses his heartfelt gratitude to the farmers who willingly let him collect samples from their farms.

## Authors' contributions

CK collected the samples and performed the RNA extractions. JDI did the HTS data analysis and submission into the appropriate repository. AG advised on the study design. All authors read and approved the final manuscript.

## Funding

The research was funded by the W.K. Kellogg Foundation Southern African Scholarship as part of Charles Karavina's Ph.D. studies at the University of KwaZulu-Natal. The funding body had no role in the experimental design, collection, analysis and interpretation of data or in writing the manuscript.

## Availability of data materials

The ZSSV genome sequences generated in this study can be freely and openly accessed on the NCBI GenBank under the Accession Numbers MK204479.1, MK204480.1 and MK204481.1. Please see Table 1 for details and links.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup> Department of Crop Science, Bindura University of Science Education, Astra Campus, P. Bag 1020, Bindura, Zimbabwe. <sup>2</sup> Discipline of Plant Pathology, University of KwaZulu-Natal, Agriculture Campus, 1 Agric Avenue, Scottsville, Pietermaritzburg 3209, Republic of South Africa.

Received: 30 September 2019 Accepted: 28 January 2020

Published online: 03 February 2020

## References

1. Weng Y, Sun Z. Major cucurbit crops. In: Wang Y-H, Behera TK, Kole C, editors. Genetics, genomics and breeding of cucurbits. Boca Raton: CRC Press; 2012. p. 1–16. <https://doi.org/10.1201/b11436>.
2. McCreight JD. Cultivation and uses of cucurbits. In: Grumet, Rebecca Katzir N, Garcia-Mas J, editors. Genetics and genomics of Cucurbitaceae. Cham: Springer International Publishing; 2016. p. 1–12.
3. Lecoq H, Desbiez C. Viruses of Cucurbit crops in the mediterranean region. An ever-changing picture. In: Loebenstein G, Lecoq H, editors. Advances in virus research. Amsterdam: Elsevier; 2012. p. 67–126. <https://doi.org/10.1016/b978-0-12-394314-9.00003-8>.
4. Lecoq H. Cucurbits. In: Loebenstein G, Thottappilly G, editors. Virus and Virus-like diseases of major crops in developing countries. Dordrecht: Springer; 2003. p. 665–88. [https://doi.org/10.1007/978-94-007-0791-7\\_26](https://doi.org/10.1007/978-94-007-0791-7_26).
5. Adams I, Fox A. Diagnosis of plant viruses using next-generation sequencing and metagenomic analysis. In: Wang A, Zhou X, editors. Current research topics in plant virology. Cham: Springer International Publishing; 2016. p. 323–35. [https://doi.org/10.1007/978-3-319-32919-2\\_14](https://doi.org/10.1007/978-3-319-32919-2_14).
6. Massart S, Chiumenti M, DeJonghe K, Glover R, Haegeman A, Koloniuk I, et al. Virus detection by high-throughput sequencing of small RNAs: large-scale performance testing of sequence analysis strategies. Phytopathology. 2019;109:488–97.
7. Massart S, Olmos A, Jijakli H, Candresse T. Current impact and future directions of high throughput sequencing in plant virus diagnostics. Virus Res. 2014;188:90–6. <https://doi.org/10.1016/J.VIRUSRES.2014.03.029>.
8. Wu Q, Ding S-W, Zhang Y, Zhu S. Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. Annu Rev Phytopathol. 2015;53:425–44. <https://doi.org/10.1146/annurev-phyto-080614-120030>.
9. Ibaba JD, Laing MD, Gubba A. First report of a novel potyvirus from the *Papaya ringspot virus* cluster infecting Zucchini (*Cucurbita pepo*) in KwaZulu-Natal, Republic of South Africa. Plant Dis. 2015;99:1289. <https://doi.org/10.1094/PDIS-02-15-0143-PDN>.
10. Ibaba JD, Laing MD, Gubba A. Zucchini shoestring virus: a distinct potyvirus in the *papaya ringspot virus* cluster. Arch Virol. 2016;161:2321–3. <https://doi.org/10.1007/s00705-016-2899-3>.
11. Adams MJ, Lefkowitz EJ, King AMQ, Harrach B, Harrison RL, Knowles NJ, et al. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). Arch Virol. 2017;162:2505–38. <https://doi.org/10.1007/s00705-017-3358-5>.
12. Wylie SJ, Adams M, Chalam C, Kreuzer J, López-Moya JJ, Ohshima K, et al. ICTV virus taxonomy profile: *Potyviriidae*. J Gen Virol. 2017;98:352–4. <https://doi.org/10.1099/jgv.0.000740>.
13. Chung BY-W, Miller WA, Atkins JF, Firth AE. An overlapping essential gene in the *Potyviriidae*. Proc Natl Acad Sci. 2008;105:5897–902. <https://doi.org/10.1073/pnas.0800468105>.

14. Olsper A, Chung BY, Atkins JF, Carr JP, Firth AE. Transcriptional slippage in the positive-sense RNA virus family Potyviridae. *EMBO Rep.* 2015;16:995–1004. <https://doi.org/10.15252/embr.201540509>.
15. Rodamilans B, Valli A, Mingot A, San León D, Baulcombe D, López-Moya JJ, et al. RNA polymerase slippage as a mechanism for the production of frameshift gene products in plant viruses of the *Potyviridae* family. *J Virol.* 2015;89:6965–7. <https://doi.org/10.1128/jvi.00337-15>.
16. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
17. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19:455–77. <https://doi.org/10.1089/cmb.2012.0021>.
18. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST: architecture and applications. *BMC Bioinform.* 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421>.
19. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947–8. <https://doi.org/10.1093/bioinformatics/btm404>.
20. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35:1547–9. <https://doi.org/10.1093/molbev/msy096>.
21. Ala-Poikela M, Goytia E, Haikonen T, Rajamaki M-L, Valkonen JPT. Helper component proteinase of the genus *Potyvirus* is an interaction partner of translation initiation factors *elf(iso)4E* and *elf4E* and contains a 4E binding motif. *J Virol.* 2011;85:6784–94.
22. Shibolet Y, Haronsky E, Leibman D, Arazi T, Wassenegger M, Whitham SA, et al. The conserved FRNK box in HC-Pro, a plant viral suppressor of gene silencing, is required for small RNA binding and mediates symptom development. *J Virol.* 2007;81:13135–48.
23. López-Moya JJ, Wang RY, Pirone TP. Context of the coat protein DAG motif affects potyvirus transmissibility by aphids. *J Gen Virol.* 1999;80:3281–8.
24. Romay G, Lecoq H, Desbiez C. *Zucchini tigré mosaic virus* is a distinct potyvirus in the *papaya ringspot virus* cluster: molecular and biological insights. *Arch Virol.* 2014;159:277–89.
25. Desbiez C, Wipf-Scheibel C, Millot P, Verdin E, Dafalla G, Lecoq H. New species in the papaya ringspot virus cluster: insights into the evolution of the PRSV lineage. *Virus Res.* 2017;241:88–94. <https://doi.org/10.1016/j.VIRUSRES.2017.06.022>.
26. Kidanemariam DB, Sukal AC, Abraham AD, Njuguna JN, Stomeo F, Dale JL, et al. Molecular characterisation of a putative new polerovirus infecting pumpkin (*Cucurbita pepo*) in Kenya. *Arch Virol.* 2019;164:1717–21.
27. Ibaba JD, Laing MD, Gubba A. Incidence and phylogeny of viruses infecting cucurbit crops in KwaZulu-Natal, Republic of South Africa. *Crop Prot.* 2015;75:46–54. <https://doi.org/10.1016/j.cropro.2015.04.019>.
28. Owolabi AT, Rabenstein F, Ehrig F, Maiss Edgar M, Vetten HJ. Strains of *Moroccan watermelon mosaic virus* isolated from *Lagenaria breviflorus* and *Coccinia barteri* in calabar, southeastern Nigeria. *Int J Virol.* 2012;8:258–70.
29. Menzel W, Abang MM, Winter S. Characterization of *Cucumber vein-clearing virus*, a whitefly (*Bemisia tabaci* G.)-transmitted carlavirus. *Arch Virol.* 2011;156:2309–11.
30. Yakoubi S, Lecoq H, Desbiez C. *Algerian watermelon mosaic virus* (AWMV): a new potyvirus species in the PRSV cluster. *Virus Genes.* 2008;37:103–9.
31. Arocha Y, Vigheri N, Nkoy-Florent B, Bakwanamaha K, Bolomphety B, Kasongo M, et al. First report of the identification of *Moroccan watermelon mosaic virus* in papaya in Democratic Republic of Congo. *Plant Pathol.* 2008;57:387. <https://doi.org/10.1111/j.1365-3059.2007.01658.x.5>.
32. Lecoq H, Dafalla G, Desbiez C, Wipf-Scheibel C, Delécolle B, Lanina T, et al. Biological and molecular characterization of *Moroccan watermelon mosaic virus* and a potyvirus isolate from Eastern Sudan. *Plant Dis.* 2001;85:547–52.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

