



CRUISE, a Tool for the Detection of Iterons in Circular Rep-Encoding Single-Stranded DNA Viruses

Adam Jones,^{a,b} George W. Kasun,^{a,c} Joel Stover,^a  Kenneth M. Stedman,^a  Ignacio de la Higuera^a

^aDepartment of Biology, Center for Life in Extreme Environments, Portland State University, Portland, Oregon, USA

^bWestview High School, Portland, Oregon, USA

^cAbSci, Vancouver, Washington, USA

ABSTRACT Iterons are short, repeated DNA sequences that are important for the replication of circular single-stranded DNA viruses. No tools that can reliably predict iterons are currently available. The CRUcivirus Iteron SEarch (CRUISE) tool is a computational tool that identifies iteron candidates near stem-loop structures in viral genomes.

Circular Rep-encoding single-stranded DNA (CRESS-DNA) viruses replicate via rolling circle replication (RCR) (1, 2). This process begins with the binding of the viral replication-associated protein (Rep) to the origin of replication in the viral double-stranded DNA (dsDNA) replicative intermediate, which is facilitated by iterons (3–5). Iterons are short sequence repeats near stem-loop structures at the genomes' origin of replication (6) (Fig. 1A). Iterons are also hypothesized to constrain recombination events in CRESS-DNA viruses (7), which makes them particularly relevant to the analysis of chimeric genomes such as those of cruciviruses (8–10). Iteron identification is essential for the analysis of CRESS-DNA virus genomes and critical for the design of experiments aimed at understanding RCR processes. However, iterons have traditionally been found manually, because there are no publicly available tools to automate the process. The CRUcivirus Iteron SEarch (CRUISE) tool performs this task by efficiently and accurately identifying and ranking iteron candidates in CRESS-DNA virus genomes.

CRUISE is written in Python 3. It uses general feature format (GFF3) sequence files for both input and output, adding iteron annotations to already annotated genomes (Fig. 1B). CRUISE performs a text search to find exact direct repeats in the region of the genome near an annotated origin of replication (stem-loop and nonnucleotide motif, as predicted with StemLoop-Finder [11]). After a preliminary filter is used to remove invalid results, an additional search looks for inexact matches to current candidate iteron sets. CRUISE then ranks the sets of iteron candidates based on metrics such as repeat length, distances from the origin of replication and between repeats, base content, and similarity to known iterons. Ambiguous bases are not accounted for, because the additional variation would produce an overwhelming number of results. CRUISE can be launched from a command-line interface, and the parameters that influence the scoring algorithm can be customized to select for preferred iteron features. CRUISE can also use a database of previously discovered iteron sequences to find matches in the input set. The tool was designed for crucivirus genome analysis but is easily adapted to any CRESS-DNA genome.

CRUISE was applied to a database of 278 recently discovered crucivirus genomes with annotated origins of replication (9), and 584 iteron candidates were found in 262 genomes. When used on the same set of genomes with an iteron sequence database, CRUISE marked an additional 223 iteron candidates. Additionally, CRUISE was applied to a set of 35 circular DNA virus genomes that had previously been manually searched for iterons (12). All manually identified iteron candidates within the set parameters were automatically identified by CRUISE. Furthermore, CRUISE identified at least 1 novel iteron set in 23 of those genomes, many of which appear to be better candidates than the previously identified iteron sets.

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2022 Jones et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ignacio de la Higuera, ide@pdx.edu.

The authors declare no conflict of interest.

Received 21 October 2022

Accepted 17 November 2022

Published 1 December 2022



FIG 1 (A) Diagram of the predicted origin of replication of CruV-249, including the stem-loop structure (with the conserved nonanucleotide motif shaded red) predicted by StemLoop-Finder (11), and 3 different iteron sets predicted by CRUISE. (B) Annotations generated by StemLoop-Finder (labeled stem-loop and nona) and by CRUISE (different iteron sets) for CruV-249, after the GFF3 output file was imported into Geneious Prime. Different iteron sets within the same genome were automatically labeled with different colors by the software.

To our knowledge, CRUISE is the only currently available iteron prediction tool. CRUISE provides a novel way to automate an important part of single-stranded DNA (ssDNA) genome annotation. This tool relies on stem-loop and nonanucleotide annotations, which can be provided by the sister program StemLoop-Finder (11). In combination, the two tools should provide researchers with a rapid and efficient way to predict CRESS-DNA virus origins of replication in genomic datasets.

Data availability. The software source code is available at a public GitHub repository (<https://github.com/adamjnes/CRUISE>). It will remain freely available for the next 10 years alongside instructions for use and any applicable updates.

ACKNOWLEDGMENTS

This work was supported by grant MCB-2025305 from the National Science Foundation, the Apprenticeships in Science and Engineering Program, Portland State University, John Howieson, and Alison Stenger.

REFERENCES

- Saunders K, Lucy A, Stanley J. 1991. DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nucleic Acids Res* 19:2325–2330. <https://doi.org/10.1093/nar/19.9.2325>.
- Ilyina TV, Koonin EV. 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Res* 20:3279–3285. <https://doi.org/10.1093/nar/20.13.3279>.
- Londoño A, Riego-Ruiz L, Argüello-Astorga GR. 2010. DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. *Arch Virol* 155:1033–1046. <https://doi.org/10.1007/s00705-010-0674-4>.
- Steinfeldt T, Finsterbusch T, Mankertz A. 2001. Rep and Rep' protein of porcine circovirus type 1 bind to the origin of replication in vitro. *Virology* 291:152–160. <https://doi.org/10.1006/viro.2001.1203>.
- Fontes EPB, Eagle PA, Sipe PS, Luckow VA, Hanley-Bowdoin L. 1994. Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *J Biol Chem* 269:8459–8465. [https://doi.org/10.1016/S0021-9258\(17\)37216-2](https://doi.org/10.1016/S0021-9258(17)37216-2).
- Argüello-Astorga GR, Guevara-González RG, Herrera-Estrella LR, Rivera-Bustamante RF. 1994. Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. *Virology* 203:90–100. <https://doi.org/10.1006/viro.1994.1458>.
- Bull SE, Briddon RW, Sserubombwe WS, Ngugi K, Markham PG, Stanley J. 2007. Infectivity, pseudorecombination and mutagenesis of Kenyan cassava mosaic begomoviruses. *J Gen Virol* 88:1624–1633. <https://doi.org/10.1099/vir.0.82662-0>.
- Diemer GS, Stedman KM. 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol Direct* 7:13. <https://doi.org/10.1186/1745-6150-7-13>.
- de la Higuera I, Kasun GW, Torrance EL, Pratt AA, Maluenda A, Colombet J, Bisseux M, Ravet V, Dayaram A, Stainton D, Kraberger S, Zawar-Reza P, Goldstien S, Briskie JV, White R, Taylor H, Gomez C, Ainley DG, Harding JS, Fontenele RS, Schreck J, Ribeiro SG, Oswald SA, Arnold JM, Enault F, Varsani A, Stedman KM. 2020. Unveiling crucivirus diversity by mining metagenomic data. *mBio* 11:e01410-20. <https://doi.org/10.1128/mBio.01410-20>.

10. Roux S, Enault F, Bronner G, Vault D, Forterre P, Krupovic M. 2013. Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. *Nat Commun* 4:2700. <https://doi.org/10.1038/ncomms3700>.
11. Pratt AA, Torrance EL, Kasun GW, Stedman KM, de la Higuera I. 2021. StemLoop-Finder: a tool for the detection of DNA hairpins with conserved motifs. *Microbiol Resour Announc* 10:e00424-21. <https://doi.org/10.1128/MRA.00424-21>.
12. Kraberger S, Argüello-Astorga GR, Greenfield LG, Galilee C, Law D, Martin DP, Varsani A. 2015. Characterisation of a diverse range of circular replication-associated protein encoding DNA viruses recovered from a sewage treatment oxidation pond. *Infect Genet Evol* 31:73–86. <https://doi.org/10.1016/j.meegid.2015.01.001>.