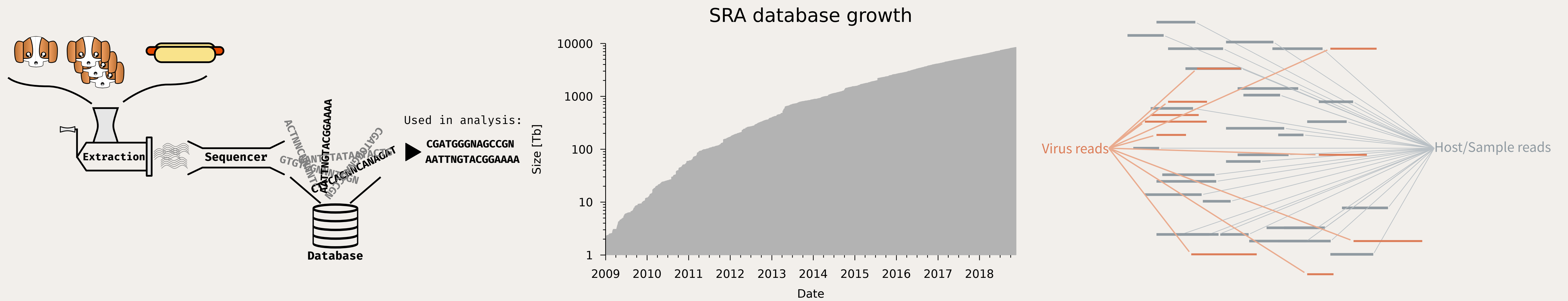


# Endovir: A pipeline to identify endogenous viruses in metagenomic data sets

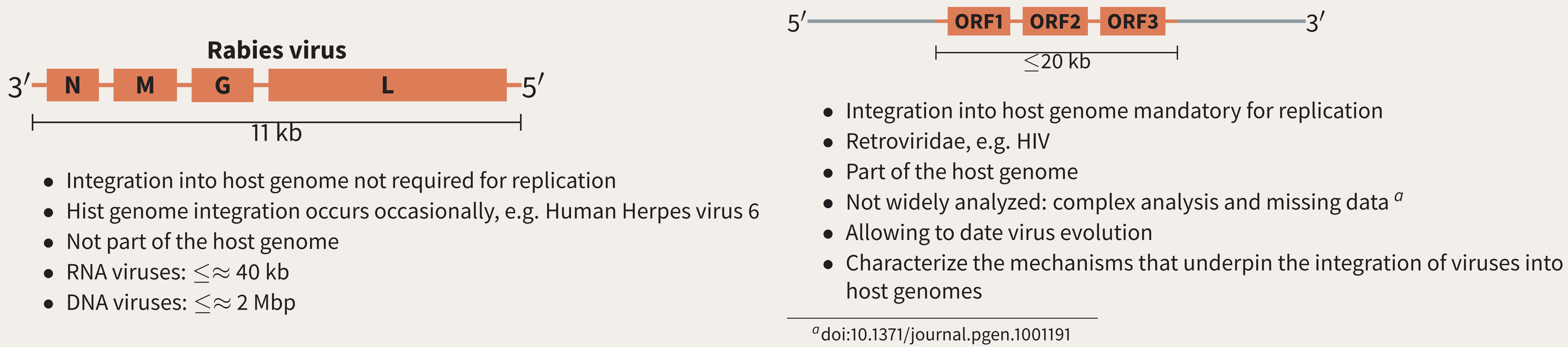
Jan P. Buchmann<sup>0</sup>, Mitchell A. Ellison II<sup>1</sup>, Jacob Waldman<sup>1</sup>, Andrew Clugston<sup>1</sup>, Paul G. Cantalupo<sup>1</sup>, Vineet Raghu<sup>1</sup>, Cody Glickman<sup>2</sup>, Kristyna Kupkova<sup>3</sup>, Karina Zile<sup>4</sup>, Ben Busby<sup>4</sup> and Edward C. Holmes<sup>0</sup>

jan.buchmann@sydney.edu.au, <https://github.com/NCBI-Hackathons/EndoVir>

## NCBI SRA data: a treasure trove for virus genomics

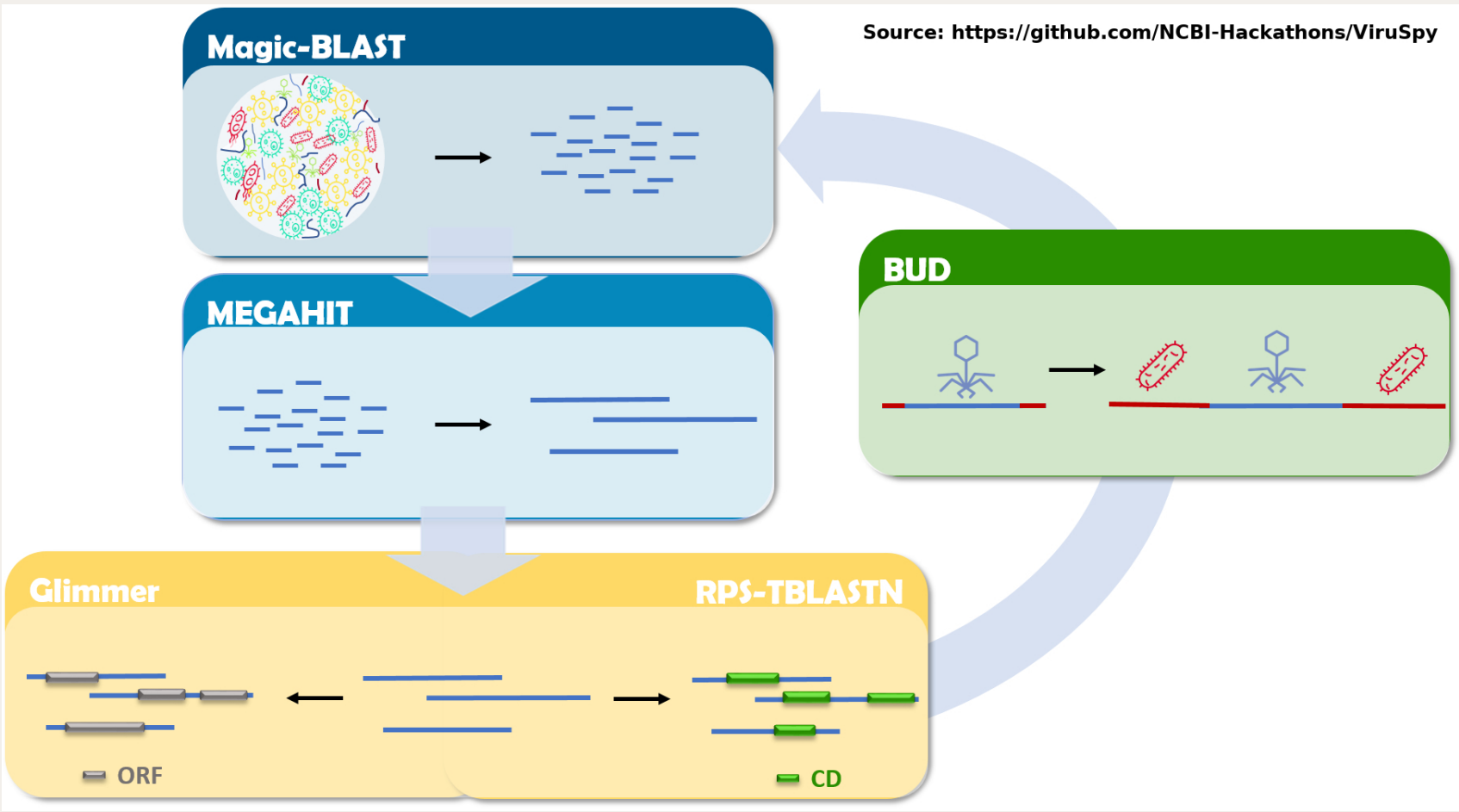


## Exogenous and Endogenous viruses

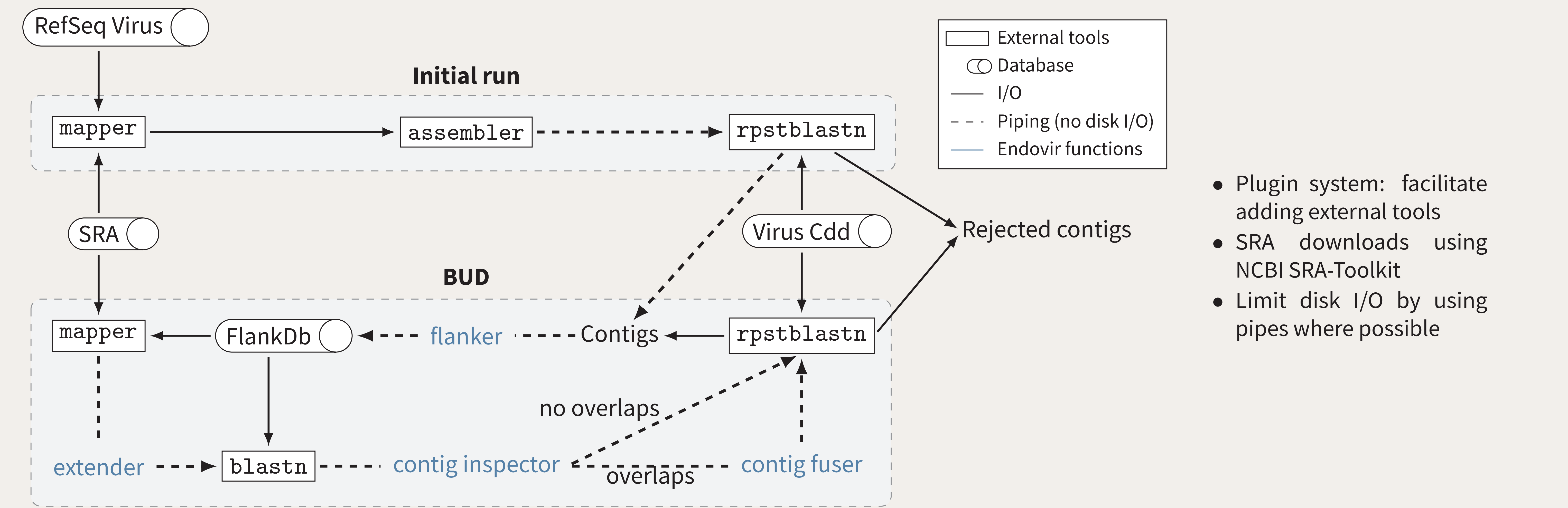


## Approach

**Collaboration** Dr. Ben Busby, NCBI/NIH  
**Built on ViruSpy** <https://github.com/NCBI-Hackathons/ViruSpy>  
**Modern** Implement in Python 3 ( $\geq 3.6$ )  
**Independent** only standard library, SRA-Toolkit, and scipy  
**Modular** Changing external tools, e.g. assembler, without touching Endovir internals  
**Status** In development, second revision



## Workflow



## Funding

- This work is supported by a National Health and Medical Research (NHMRC) Australia Fellowship awarded to Edward C. Holmes.
- Initial development was funded by an ORISE scientific visitor fellowship to Jan P. Buchmann

## Affiliations

- The University of Sydney, Sydney, Australia
- The University of Pittsburgh, Pittsburgh, PA 15224, USA
- The University of Colorado, Denver, Aurora, Colorado 80045, USA
- Brno University of Technology, Brno, 612 00, Czech Republic
- NCBI, NLM, NIH, Bethesda, MD 20894, USA