Briefings in Bioinformatics, 2026, 27, bbag317
https://doi.org/10.1093/bib/bbag317
Published: 17 June 2026
Letter to Editor

OXFORD

Response to 'On using clustering statistics for assessing plasmid binning tools accuracy'

Marco Teixeira[1,2], Colin J. Worby[1], Thomas Abeel[1,2], Ashlee M. Earl[1], Abigail L. Manson[1,*]

[1]Infectious Disease and Microbiome Program, Broad Institute of Massachusetts Institute of Technology and Harvard, 415 Main St, Cambridge, MA 02142, United States
[2]Delft Bioinformatics Lab, Department of Intelligent Systems, Delft University of Technology, Van Mourik Broekmanweg 6, Delft 2628XE, Zuid Holland, The Netherlands
*Corresponding author. Infectious Disease and Microbiome Program, Broad Institute of Massachusetts Institute of Technology and Harvard, 415 Main St, Cambridge, MA 02142, United States. E-mail: amanson@broadinstitute.org

Abstract

This response addresses the comments raised by Dr. Epain and colleagues in their Letter to the Editor titled 'On using clustering statistics for assessing plasmid binning tools accuracy' in response to our paper 'Circling in on plasmids: benchmarking plasmid detection and reconstruction tools for short-read data from diverse species'. In their letter, the authors caution against using homogeneity and completeness measures to evaluate the accuracy of plasmid binning tools due to issues related to the length and content of contigs. They also refer to *PlasEval*, a plasmid binning evaluation tool that they recently developed. In response, we evaluated the impact of contig size and content on the results of our study, and also repeated the benchmarking of plasmid reconstruction tools using *PlasEval*. Though the absolute values of metrics differed across tests, we observed nearly identical rankings among plasmid reconstruction tools as originally reported in our study, with gplas2 outperforming all other tools. Nevertheless, approaches specifically designed for plasmid data may be better suited than clustering metrics to evaluate plasmid reconstruction.

Keywords **plasmid reconstruction, plasmid, bacterial genomics, genomics**

Main

We thank Dr. Epain and colleagues for interest in our publication [1], and for raising important points that gave us the opportunity to explore the impact of our choice of metrics on the performance rankings of various plasmid reconstruction tools.

In their letter, Epain *et al.* claimed that the well-established clustering metrics [2, 3] that we adopted for our study – homogeneity, completeness, and normalized mutual information (NMI) – were not well-suited to evaluate the accuracy of plasmid reconstruction. First, the authors suggested that abundant *short contigs* can skew clustering metrics by giving equal weight to all contigs regardless of size. Second, they argued that *chromosomal contigs* are easier to group, and that including them distorts clustering metrics. Third, the authors pointed out that adapting clustering metrics to allow for *repeat contigs* – contigs that are present in multiple plasmids – could lead to artificially low scores. Finally, the authors highlighted *PlasEval*, a tool that they recently developed to evaluate plasmid reconstruction that addresses the aforementioned limitations [4].

Our choice of metrics was motivated by the specific goals of our benchmarking: to evaluate plasmid reconstruction for less well-characterized organisms, with fewer plasmids in the databases used by tools, and to investigate the impact of antimicrobial resistance genes (ARGs) on plasmid reconstruction. Nevertheless, to further assess our choice of metrics, including their impact on the results of our published study, we compared the clustering metrics in our original work to modified versions of these metrics that account for the limitations pointed out by Epain *et al.*, as well as to metrics from *PlasEval.*

Short contigs. We repeated our benchmarking using modified metrics for homogeneity, completeness, and NMI, weighted by the length of each contig (Supplementary Fig. 1). Though the absolute values of these metrics varied from our original measures, they did not change the ranking of tools relative to our original work. Notably, nearly 50% of contigs (n = 68) with ARGs in our dataset were short (<5000 bp long), likely due to their proximity to mobile genetic elements that induce breaks in the assembly [5, 6]. Given the critical importance of ARGs in plasmid research, particularly for the organisms included in our benchmarking, we chose to equally weight contigs, regardless of size.

Chromosomal contigs. We repeated our benchmarking removing unbinned chromosomal contigs. As the authors suggested, the absolute values of metrics decreased. However, the relative performance of tools was similar to that of our original work (Supplementary Fig. 1).

Received: May 11, 2026. **Accepted:** May 21, 2026

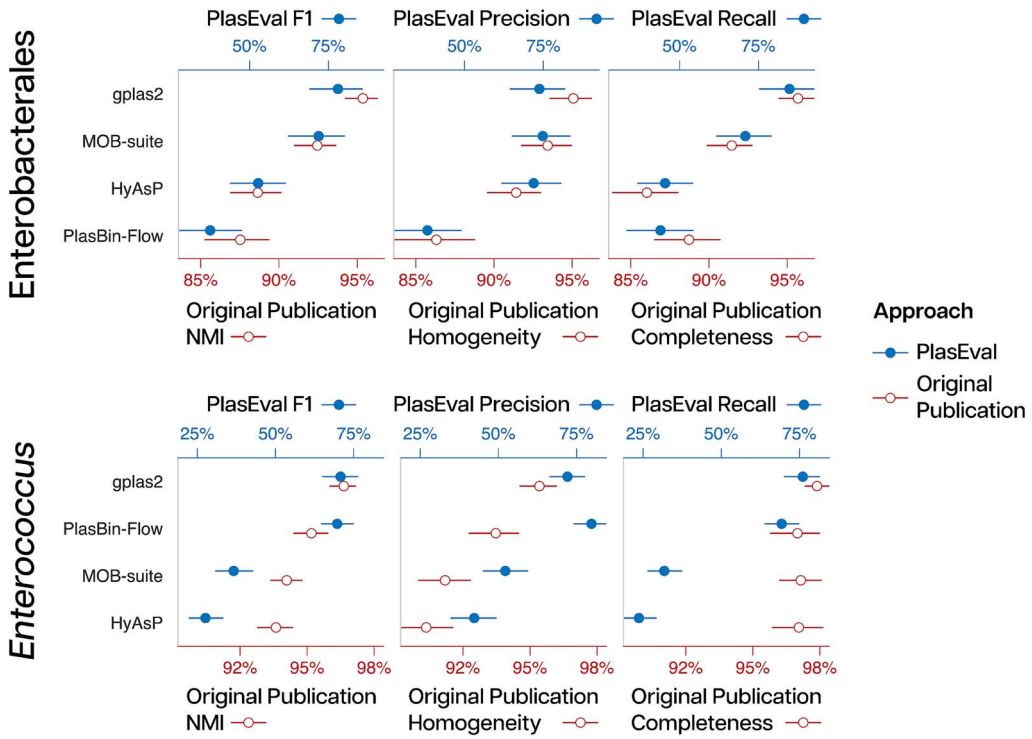

Figure 1 Plasmid reconstruction metrics in our original publication (open markers, bottom x-axis) and from PlasEval (filled markers, top x-axis) per tool for Enterobacterales and enterococcus, with bars representing the 95% confidence interval.

Repeat contigs. We evaluated the impact of repeat contigs by benchmarking plasmid reconstruction tools using *PlasEval*, which was not available at the time we performed our analysis. We found that the relative performance of plasmid reconstruction tools was similar regardless of which evaluation metrics were used (Fig. 1). Gplas2 outperformed all other tools both for Enterobacterales and enterococci, and the relative ranking of tools remained unchanged. However, we observed a relative improvement in PlasBin-Flow's performance for enterococci when using *PlasEval* as compared to our original metrics. This might have been a result of the larger number of enterococcal contigs assigned to multiple plasmids by PlasBin-Flow (1.0%) relative to the other tools (<0.2%). Therefore, *PlasEval* may be better suited for assemblies with many repeat contigs and when evaluating tools more susceptible to assigning contigs to multiple plasmid bins.

In conclusion, we observed largely similar results, including a similar ranking of tools when comparing our original clustering metrics with those accounting for contig size, chromosomal contigs, and repeats (including *PlasEval*) to evaluate plasmid reconstruction accuracy. The highest-scoring tool was the same, regardless of the choice of metric. Nevertheless, approaches specifically designed for plasmids, such as *PlasEval*, may be better suited than clustering metrics to evaluate plasmid reconstruction.

Key Points

- We repeated the benchmarking of plasmid reconstruction tools using the clustering metrics in our original work, adapted versions of these metrics accounting for contig length and chromosomal contigs, and *PlasEval*.
- The ranking of tools was nearly identical across metrics, with gplas2 performing the best.
- Nevertheless, approaches specifically designed for plasmid data may be better suited than clustering metrics to evaluate plasmid reconstruction.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Funding

This work was supported by Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Grant Numbers U19AI110818 to the Broad Institute and U19AI172725 to University of California Irvine and the Broad Institute.

Data availability

The code used for analysis is available at https://github.com/broadinstitute/plasmid-detection-benchmark.

References

1. Teixeira M, Souque C, Worby CJ *et al.* Circling in on plasmids: benchmarking plasmid detection and reconstruction tools for short-read

data from diverse species. *Brief Bioinform* 2025;**26**:bbaf589. https://doi.org/10.1093/bib/bbaf589

2. Nguyen R, Sokhansanj BA, Polikar R *et al.* Complet+: a computationally scalable method to improve completeness of large-scale protein sequence clustering. *PeerJ* 2023;**11**:e14779. https://doi.org/10.7717/peerj.14779
3. Lossio-Ventura JA, Gonzales S, Morzan J *et al.* Evaluation of clustering and topic modeling methods over health-related tweets and emails. *Artif Intell Med* 2021;**117**:102096. https://doi.org/10.1016/j.artmed.2021.102096
4. Mane A, Sanderson H, White AP *et al.* Plaseval: a framework for comparing and evaluating plasmid detection tools. *BMC Bioinform* 2024;**25**:365. https://doi.org/10.1186/s12859-024-05941-0
5. Wang Y, Dagan T. The evolution of antibiotic resistance islands occurs within the framework of plasmid lineages. *Nat Commun* 2024;**15**:4555. https://doi.org/10.1038/s41467-024-48352-8
6. Wein T, Wang Y, Hülter NF *et al.* Antibiotics interfere with the evolution of plasmid stability. *Curr Biol* 2020;**30**:3841–3847.e4. https://doi.org/10.1016/j.cub.2020.07.019

Briefings in Bioinformatics, 2026, **27**, bbag317
https://doi.org/10.1093/bib/bbag317
Letter to Editor