Briefings in Bioinformatics, 2026, 27, bbag240
https://doi.org/10.1093/bib/bbag240
Published: 19 May 2026
Letter to Editor

OXFORD

On using clustering statistics for assessing plasmid binning tools accuracy

Victor Epain, Aniket Mane, Cedric Chauve*

Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, V5A 1S6 BC, Canada

*Corresponding author. Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, V5A 1S6 BC, Canada. E-mail: cedric.chauve@sfu.ca

Abstract

We discuss the drawbacks of using homogeneity and completeness, statistics defined for the evaluation of clustering accuracy, for evaluating plasmid binning tools accuracy, motivated by the recently published paper "Circling in on plasmids: benchmarking plasmid detection and reconstruction tools for short-read data from diverse species," by Teixeira *et al*.

Keywords plasmids detection, clustering, accuracy measure

The recent paper "Circling in on plasmids: benchmarking plasmid detection and reconstruction tools for short-read data from diverse species" [1] benchmarks tools for plasmid binning, a problem motivated by applications in epidemiological surveillance among others. Plasmid binning tools group contigs from draft assemblies into *plasmid bins* with the aim that each bin contains the set of contigs present in a plasmid of the assembled genome. To assess the accuracy of plasmid binning tools, the authors use entropy-based clustering accuracy measures, *completeness* and *homogeneity* [2, 3]. Given a predicted clustering and a ground truth clustering of the same ground set, homogeneity quantifies the *purity*, i.e. the extent to which a predicted cluster contains elements from several true clusters, while completeness quantifies the *fragmentation*, i.e. the extent to which a true cluster is fragmented in the predicted clusters. Both measures are used to compare the bins generated by plasmid binning tools (predicted bins) against ground truth bins (true bins) obtained from long-reads sequencing data. In this letter, we caution that using homogeneity and completeness to evaluate plasmid binning accuracy suffers from drawbacks, which we outline and illustrate below.

Contigs length. Homogeneity and completeness do ignore the fact that contigs represent sequences of various lengths, which gives more weight to short contigs, that are often numerous in draft assemblies.

Non-plasmid contigs. Generally predicted and true bins do not contain the same sets of contigs, with some true plasmid contigs missing in the predicted bins and some predicted plasmid contigs actually being chromosomal contigs. Teixeira *et al.* [1] assign, for a given set of bins (predicted or true), all contigs that do not appear in any bin to a (predicted or true) *chromosome bin*. The predicted and true chromosome bins are generally much larger than the sets of plasmid bins and highly similar to each other, which distorts the homogeneity and completeness toward values that overestimates the actual accuracy of the predicted bins.

Repeat contigs. For the purpose of plasmid binning, a *repeat* is a contig, i.e. shared across multiple (predicted or true) bins; it is well known that plasmid contain repeated sequences that result in such repeats, e.g. Insertion Sequences. So plasmid binning is not a pure clustering problem on a ground set defined as the contigs of a draft assembly, and resolving which copy of a repeat in the predicted bins matches with which copy in the true bins is essential to quantify the accuracy. Teixeira *et al.* [1] address this issue by taking the cross-product of repeats between predicted and true bins, and consider each resulting pair as a *pseudo-contig* in the clustering. This leads to the undesirable property that, in the presence of repeats, perfect predicted bins (i.e. identical to the true bins) will not exhibit perfect homogeneity and completeness. For example, if a repeat is present in true bins T_1 and T_2 as well as in predicted bins P_1 and P_2, the pseudo-contigs are the pairs $(T_1, P_1), (T_2, P_1), (T_1, P_2)$, and (T_2, P_2), and at most two pseudo-contigs are considered as correct, while at least two pseudo-contigs are considered as erroneous, decreasing both homogeneity and completeness.

Illustration. We illustrate the issues described above with two experiments.

We illustrate the impact of the chromosomal bin on 48 samples considered in [1]. We replaced the chromosome bin as defined in [1] by a *reduced chromosome bin* excluding all contigs that appear in neither a predicted bin nor a true bin, and computed the corresponding homogeneity and completeness. In this experiment, we account for the length of contigs when computing homogeneity and

Received: March 24, 2026. **Revised:** March 24, 2026. **Accepted:** April 22, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

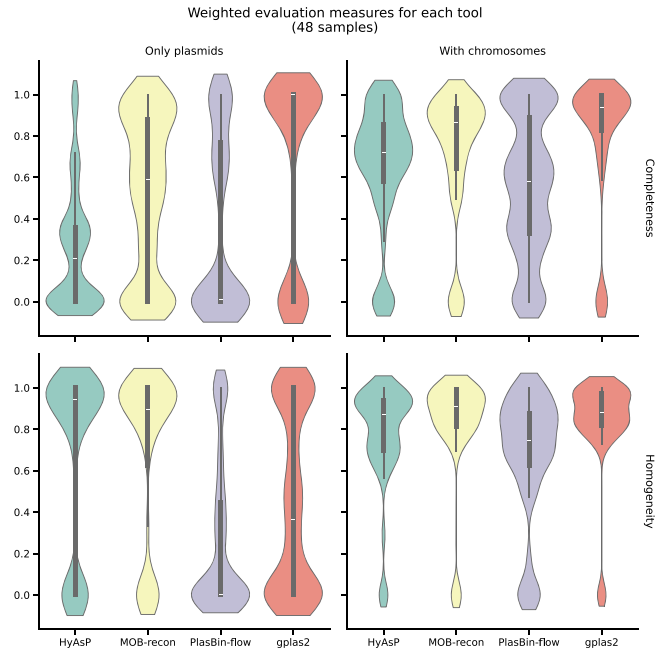

Figure 1 Homogeneity and completeness distributions on 48 sample, for 4 plasmid binning tools. Top left: completeness with a reduced chromosome bin. Top right: completeness with a full chromosome bin. Bottom left: homogeneity with a reduced chromosome bin. Bottom right: homogeneity with a full chromosome bin.

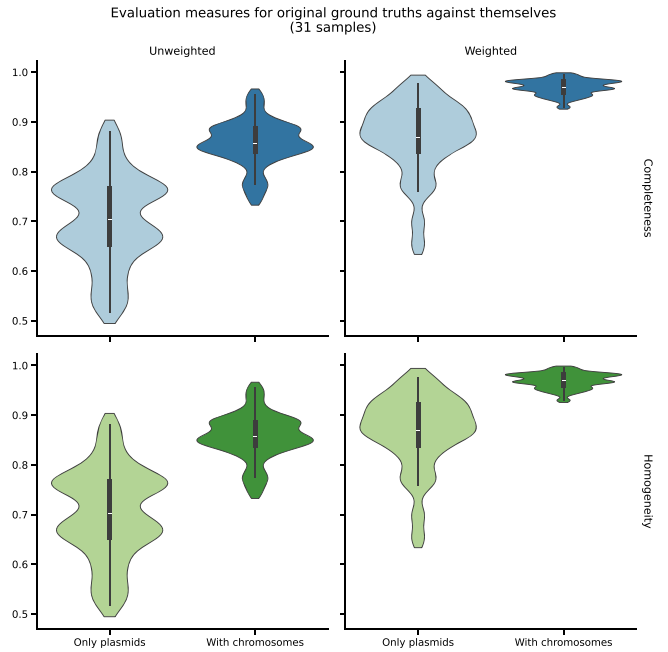

Figure 2 Assessing purity and fragmentation with perfect predicted bins when using the approach of Teixeira *et al*. [1] to handle repeats. Top left: unweighted completeness without (left) and with (right) a chromosome bin. Top right: weighted completeness without (left) and with (right) a chromosome bin. Bottom left: unweighted homogeneity without (left) and with (right) a chromosome bin. Top right: weighted homogeneity without (left) and with (right) a chromosome bin.

completeness. Figure 1 shows that using the full chromosome bin biases the homogeneity and completeness toward higher values and can lead to different conclusions as to the respective accuracy of the evaluated tools.

To illustrate the impact of repeats, we considered the set of 31 samples for which the true bins contain repeat(s) and used the true bins as perfect predicted bins, so ensuring by construction perfect purity (respectively, fragmentation) of the predicted (respectively, true) bins. We provide measures which do (respectively, do not) account for the length of contigs, weighted (respectively, unweighted) measures. Figure 2 shows that accounting for repeats as in [1] results in homogeneity and completeness that can be far from 1.

Conclusion. While many bioinformatics problems can be seen as clustering problems, they often differ from pure clustering by specific elements that need to be accounted for in adapting or designing accuracy statistics. For assembly-type problems, such as plasmid binning, handling repeats is often challenging both in designing analysis methods and in designing appropriate accuracy measures. To illustrate this, we refer to PlasEval [4], a measure of accuracy for plasmid binning that handles repeats by seeking an optimal matching between repeats in predicted and true bins. This can be computationally costly but ensures that perfect predictions result in perfect purity and fragmentation scores. This illustrates the challenge of the balance between the computational footprint required for computing sound accuracy measures and the need to satisfy desirable statistical properties.

Key Points

- Plasmid binning bears strong similarity to clustering problems.
- Standard clustering statistics require to be adapted to assess plasmid binning accuracy.
- Accounting for repeats in measuring plasmid binning accuracy is challenging.

Author contributions

C.C. designed the study. V.E. and A.M. performed implementation and data analysis. V.E., A.M., and C.C. wrote the manuscript.

Conflicts of interest

None declared.

Funding

None declared.

Data availability

Code and data to reproduce our experiments are available at https://github.com/vepain/lte-bib-bbaf589-10-1093.

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. *Briefings Bioinform* 2025;**26**:bbaf589.
2. Rosenberg A, Hirschberg J. V-measure: a conditional entropy-based external cluster evaluation measure. In: *EMNLP-CoNLL 2007, Proceedings of the 2007 Joint Conference on Empirical Methods in*

Natural Language Processing and Computational Natural Language Learning, June 28–30, 2007, Prague, Czech Republic. Eisner J (ed.), pp. 410–20. ACL, 2007.

3. Utt J, Springorum S, Köper M *et al.* Fuzzy v-measure - an evaluation method for cluster analyses of ambiguous data. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26–31, 2014*. Calzolari N, Choukri K, Declerck T *et al.* (eds.), pp. 581–7. European Language Resources Association (ELRA), 2014.
4. Mane AC, Sanderson H, White AP *et al.* Plaseval: a framework for comparing and evaluating plasmid detection tools. *BMC Bioinformatics* 2024;**25**:365.

© The Author(s) 2026. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.
Briefings in Bioinformatics, 2026, **27**, bbag240
https://doi.org/10.1093/bib/bbag240
Letter to Editor