

COMMENTARY

Evaluation of genomic contamination detection tools: Assessing the influence of horizontal gene transfer on their efficiency through contamination simulations at various taxonomic ranks

Tyson Mende*

Department of Ecology, University of Liège, 4000 Liège, Belgium

*Corresponding author E-mail: mende@tyson.edu

Received: 04 March, 2024, Manuscript No: UJE-24-131859; **Editor assigned:** 06 March, 2024, PreQC No: P-131859; **Reviewed:** 18 March, 2024, QC No: Q-131859; **Revised:** 23 March, 2024, Manuscript No: R-131859; **Published:** 30 March, 2024

Genomic contamination poses a significant challenge in bioinformatics and genomics research, potentially skewing results and leading to erroneous conclusions. Various tools have been developed to detect contamination, but their efficacy can be influenced by factors such as Horizontal Gene Transfer (HGT) and taxonomic ranks. In this article, we evaluate genomic contamination detection tools and investigate how HGT impacts their efficiency through contamination simulations across different taxonomic ranks. Our findings shed light on the complexities of contamination detection and offer insights into improving these tools for more accurate genomic analyses.

Keywords: Genomic contamination, Horizontal gene transfer, Contamination detection tools, Taxonomic ranks, Simulation study, Genomic research, Data quality, Reliability assessment.

Introduction

Genomic contamination, the unintentional inclusion of foreign genetic material in sequencing data, can compromise the integrity of genomic analyses. It can arise from a variety of sources, including laboratory artifacts, sample mix-ups and environmental contamination. Detecting and mitigating contamination is crucial for ensuring the reliability of genomic studies. Over the years, several computational tools have been developed to identify and remove contaminated sequences (Cornet, L., et al., 2022). However, the efficacy of these tools can vary, influenced by factors such as the level of contamination, the taxonomic origin of the contaminant and the presence of Horizontal Gene Transfer (HGT) events. Various software tools have been developed to detect genomic contamination, employing different algorithms and approaches. Some of the commonly used tools include Kraken, BlobTools, ContamFinder and DeconSeq. These tools utilize techniques such as sequence alignment, k-mer analysis and reference-based mapping to identify potential contaminants in sequencing data.

Description

BlobTools integrates taxonomic information with sequence coverage data to identify potential contaminants based on their taxonomic origin. ContamFinder utilizes statistical methods to detect contamination by comparing the composition of sequences with expected patterns. DeconSeq employs a reference-based approach to identify and remove contaminant sequences from genomic datasets (Schierwater, B., et al., 2009). While these tools offer valuable capabilities for contamination detection, their performance

can be influenced by factors such as the level of contamination, the taxonomic diversity of the dataset and the presence of HGT events.

Horizontal Gene Transfer (HGT), the transfer of genetic material between different organisms, can complicate contamination detection efforts. HGT events can introduce foreign sequences into genomes, blurring the boundaries between genuine genomic content and contaminants. Traditional contamination detection tools may struggle to differentiate between endogenous sequences and horizontally transferred elements, leading to false positives or negatives (Philippe, H., et al., 2011). The impact of HGT on contamination detection efficiency can vary depending on the taxonomic ranks involved. For example, in closely related species or strains, HGT events may involve genetic exchange between organisms with similar genomic compositions, making it challenging to distinguish between native and foreign sequences. Conversely, in distantly related taxa, HGT events may result in the acquisition of sequences that exhibit significant divergence from the host genome, facilitating their detection as potential contaminants (Laurin-Lemay, S., et al., 2012).

To evaluate the influence of HGT on contamination detection, we conducted contamination simulations using synthetic datasets representing different taxonomic ranks and varying degrees of HGT activity. By systematically introducing contaminant sequences derived from different taxonomic groups and assessing the performance of contamination detection tools, we aimed to elucidate how HGT impacts the accuracy and reliability of these tools. In our contamination simulations, we generated synthetic datasets representing microbial communities at different taxonomic ranks, including species, genus and phylum levels. We introduced varying levels of contamination, ranging from low to high, by randomly selecting sequences from unrelated taxa and integrating them into the datasets (Lupo, V., et al., 2021). We then applied a selection of contamination detection tools to the simulated datasets and evaluated their performance in identifying and removing contaminant sequences. Metrics such as sensitivity, specificity and false discovery rate were used to assess the accuracy and robustness of the tools under different conditions.

In scenarios involving closely related taxa, where HGT is prevalent, traditional tools exhibited reduced sensitivity and specificity, leading to higher false positive and false negative rates. Conversely, in scenarios with distantly related taxa and limited HGT activity, the performance of contamination detection tools improved, with higher accuracy in identifying contaminants. Genomic contamination poses a significant challenge in genomic research, with the potential to undermine the reliability of data analysis. While various tools have been developed to detect and mitigate contamination, their performance can be influenced by factors such as taxonomic diversity and the presence of Horizontal Gene Transfer (HGT) events.

Conclusion

Our evaluation of contamination detection tools and contamination simulations across different taxonomic ranks shed light on the complexities of contamination detection and the impact of HGT on tool efficiency. We observed that traditional tools may struggle to accurately detect contamination in scenarios involving HGT, particularly when dealing with closely related taxa. However, in cases with limited HGT activity and greater taxonomic divergence, the performance of contamination detection tools improves. Moving forward, there is a need to develop more robust and versatile tools capable of effectively detecting contamination across diverse taxonomic groups and accounting for the influence of HGT (Parks, D. H., et al., 2015). Integration of advanced algorithms, machine learning techniques and comprehensive reference databases may enhance the accuracy and reliability of contamination detection in genomic studies, ultimately advancing our understanding of biological systems and improving the quality of genomic data analysis.

Acknowledgement

None.

Conflict of Interest


The authors declare no conflict of interest.

References

- Cornet, L., Baurain, D. (2022). Contamination detection in genomic data: More is not enough. *Genome Biology* 23:60.
- Schierwater, B., Eitel, M., Jakob, W., Osigus, H. J., Hadrys, H., Dellaporta, S. L., DeSalle, R. (2009). Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biology* 7:e1000020.
- Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., Baurain, D. (2011). Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology* 9:e1000602.
- Laurin-Lemay, S., Brinkmann, H., Philippe, H. (2012). Origin of land plants revisited in the light of sequence contamination and missing data. *Current Biology* 22:R593-R594.
- Lupo, V., Van Vlierberghe, M., Vanderschuren, H., Kerff, F., Baurain, D., Cornet, L. (2021). Contamination in reference sequence databases: Time for divide-and-rule tactics. *Frontiers in Microbiology* 12:755101.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells and metagenomes. *Genome Research* 25:1043-1055.

Citation:

Mende, T. (2024). Evaluation of genomic contamination detection tools: Assessing the influence of horizontal gene transfer on their efficiency through contamination simulations at various taxonomic ranks. *Ukrainian Journal of Ecology*. 14:22-24.

 This work is licensed under a Creative Commons Attribution 4.0 License
