

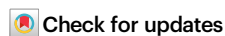
Reply to: Genome contamination may lead to an overestimation of horizontal gene transfer inferences

Received: 6 March 2026

Morten O. A. Sommer  & Christian Munck

Accepted: 22 April 2026

Published online: 19 May 2026

REPLYING TO N. Godron et al. *Nature Communications* <https://doi.org/10.1038/s41467-026-69064-1> (2026)

We thank Godron et al. for their interest in our previous article¹ and their detailed *Matters Arising* analysis of potential genome contamination². Godron et al.² argue that many of the putative horizontal gene transfer (HGT) events we identified using Sequence Read Archive (SRA) data were actually artifacts of contamination. They report that 30 of the 34 transfer events identified in our confirmatory analysis likely arose from contaminant sequences rather than true host acquisitions. For example, in one *Staphylococcaceae* isolate (SRA: ERR212931) where we reported a *catI* resistance gene alongside the IS1 mobile element, their re-assembly indicated that the contig was derived from *Acinetobacter* contaminant DNA.

Our original pipeline for confirmation of predicted potential HGT events was designed to be sensitive to detecting evidence of predicted antibiotic resistance gene (ARG) transfers, but it did not include contamination filtering based on the contig coverage. The analysis by Godron et al. has uncovered instances (particularly for *Staphylococcaceae* and *Streptococcaceae*) in which what appears to be contaminant DNA have introduced false-positive signals, inflating the apparent HGT confirmation rate in our study. However, the presence of contamination in the SRA does not invalidate the central premise of our confirmation step: that some predicted ARG transfers are present in existing genomic data. After excluding the 30 potentially contaminated or artifactual HGT cases, Godron et al. identified four HGT events (*erm*(F), *erm*(T), *tet*(H), and *tet*(W)) in *Enterobacteriaceae* that withstand scrutiny as genuine horizontal transfers. While these four HGT support the validity of our approach in uncovering real ARG dissemination events, we recognize that contamination likely has inflated the raw number of confirmed events in our original publication.

The *Matters Arising* authors suggest that contamination may have affected our assessment of mobile genetic element (MGE) host ranges based on reference genomes (RefSeq database), yet provide limited data to support the claim. In our original analysis, we curated

the RefSeq database entries to limit contamination by clustering the 16S rRNA genes from each RefSeq entry (as described in our article¹). Godron et al. investigated RefSeq entries associated with the IS1 transposase element, which our study had identified as a major driver of inter-phyla ARG transfer. Their re-analysis found that among 19 draft genomes in which our original analysis detected IS1 in a new host phylum, 7 appeared to have IS1 on contigs with differential or no coverage from the target species. While it is hard to draw any conclusions from this analysis, it underscores the need for careful contamination checks even when performing genomic analysis on the RefSeq genomes, as also highlighted by Lupo et al.³. NCBI is increasingly aware of contamination challenges and is developing new approaches to filter out contaminants, as described by Astashyn et al.⁴.

With this additional analysis in mind, future bioinformatics studies of resistance gene dissemination using our approach should include stringent contamination control to increase confidence in validating predicted HGT events, particularly when analyzing SRA entries.

References

1. Ellabaan, M. M. H. et al. Forecasting the dissemination of antibiotic resistance genes across bacterial genomes. *Nat. Commun.* **12**, 2435 (2021).
2. Godron, N., Ruppé, E. & Leclercq, S. O. Genome contamination may lead to an overestimation of horizontal gene transfer inferences. *Nat. Commun.* **17**, 1219 (2026). 1. Ellabaan M. et al. Forecasting the dissemination of antibiotic resistance genes across bacterial genomes. *Nat. Commun.* **12**, 2435 (2021).
3. Lupo, V. et al. Contamination in Reference Sequence Databases: Time for Divide-and-Rule Tactics. *Front. Microbiol.* **12**, 755101 (2021).

4. Astashyn, A. et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol.* **25**, 60 (2024).

Author contributions

Both authors wrote and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Morten O. A. Sommer.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026