



Questioning the Quality of 16S rRNA Gene Sequences Derived From Human Gut Metagenome-Assembled Genomes

Pranvera Hiseni^{1,2*}, Lars Snipen¹, Robert C. Wilson³, Kari Furu² and Knut Rudi^{1,3}

¹ Department of Chemistry, Biotechnology and Food Sciences, Norwegian University of Life Sciences, Ås, Norway, ² Genetic Analysis AS, Oslo, Norway, ³ Department of Biotechnology, Faculty of Applied Ecology, Agricultural Sciences and Biotechnology, Inland Norway University of Applied Sciences, Hamar, Norway

Keywords: 16S rRNA, metagenome assembled genome (MAG), metagenome analyses, human gut microbiome, prokaryotic genome

OPEN ACCESS

Edited and reviewed by:

Franck Carbonero,
Washington State University Health
Sciences Spokane, United States

*Correspondence:

Pranvera Hiseni
ph@genetic-analysis.com

Specialty section:

This article was submitted to
Microorganisms in Vertebrate
Digestive Systems,
a section of the journal
Frontiers in Microbiology

Received: 25 November 2021

Accepted: 28 December 2021

Published: 04 February 2022

Citation:

Hiseni P, Snipen L, Wilson RC, Furu K
and Rudi K (2022) Questioning the
Quality of 16S rRNA Gene Sequences
Derived From Human Gut
Metagenome-Assembled Genomes.
Front. Microbiol. 12:822301.
doi: 10.3389/fmicb.2021.822301

The recent introduction of metagenome-assembled genomes (MAGs) has marked a major milestone in the human gut microbiome field (Almeida et al., 2019; Nayfach et al., 2019; Pasolli et al., 2019). Such reference-free, *de novo*-assembled genomes (Hugerth et al., 2015) have revealed a wide range of hitherto uncultured microbial species in human gut samples.

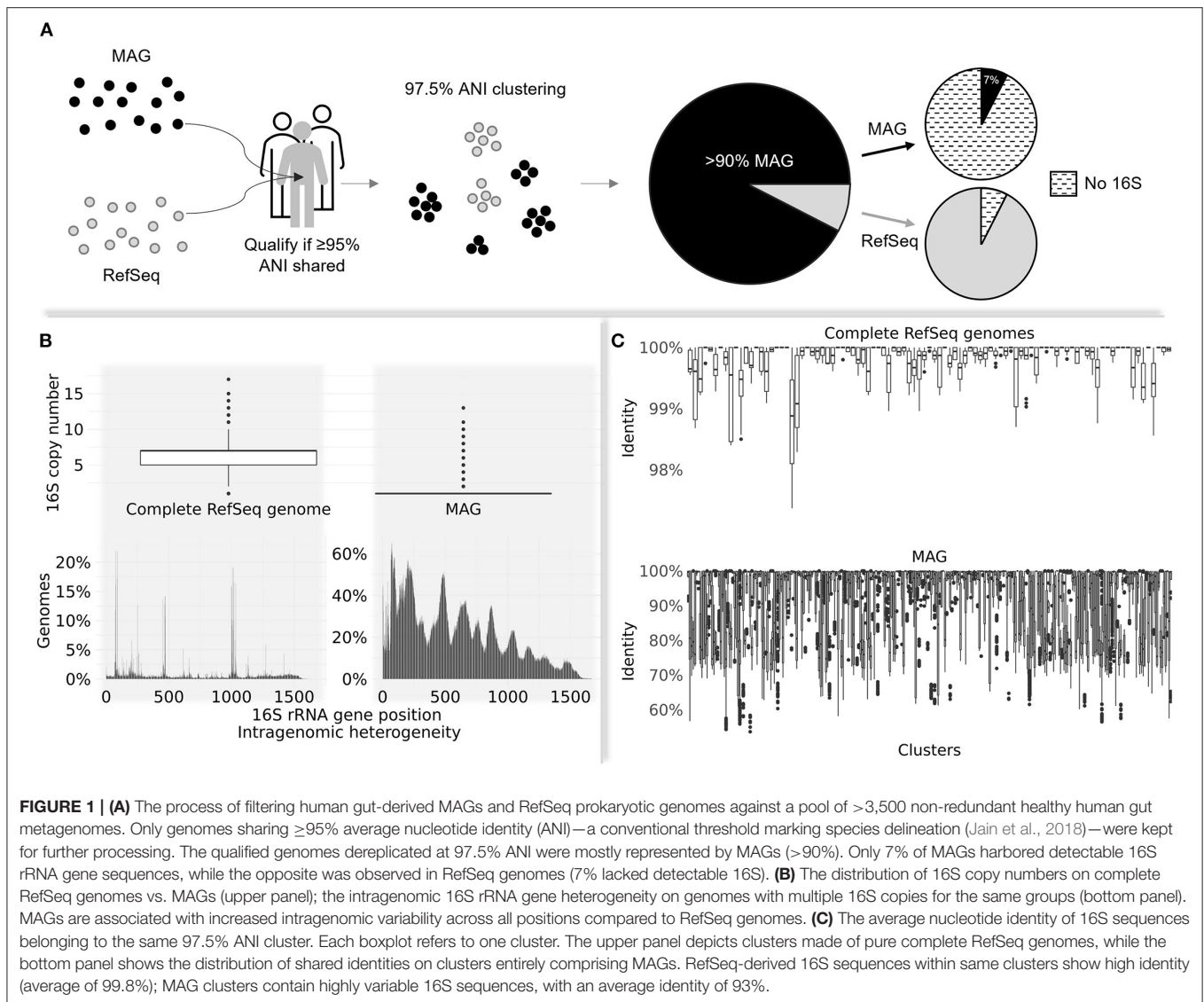
The significance of MAGs in unraveling human gut microbial diversity was supported by their overwhelming representation in a comprehensive human gut prokaryotic collection filtered by metagenome data dereplicated at 97.5% average nucleotide identity (ANI) (Hiseni et al., 2021). More than 90% of the collection consists of MAGs, while the rest of the collection mainly comprises RefSeq genomes (Figure 1A).

A great challenge related to MAGs is their lack of 16S rRNA sequences. Skewed species abundance, high 16S sequence similarity, and high volumes of short-reads data cause major difficulties for assembling the sequences of this gene (Yuan et al., 2015), frequently rendering these genomes incomplete.

A barrnap search (<https://github.com/tseemann/barrnap>) revealed that from >270,000 qualified MAGs, only 7% yielded 16S sequences, while this gene was found in 93% of >106,000 other genome types. MAGs positive for 16S had a significantly lower copy number compared to complete RefSeq genomes (Figure 1B; top panel) and substantially higher intragenomic variance (Figure 1B; bottom panel). Challenges in obtaining multiple 16S copies from incomplete genomes are well-described in the literature (Perisin et al., 2016; Louca et al., 2018); however, to exacerbate the problem, their enormous intragenomic heterogeneity renders their overall quality questionable.

A multiple sequence alignment of 16S rDNA sequences extracted from members of identical 97.5% ANI clusters, followed by the computation of their distance [*ape* package in RStudio (Paradis and Schliep, 2018)], has revealed that clusters consisting purely of MAGs share on average 93% identity, as contrasted by 99.8% average 16S sequence identity in clusters made of pure, complete RefSeq genomes (Figure 1C).

Considering that 16S is a highly conserved gene, its identity among same-cluster genomes was expected to be higher than the threshold used for dereplicating them (>97.5%; Kim et al., 2014; Jain et al., 2018). The excessive 16S divergence among MAG-only clusters raises red flags, potentially reflecting issues related to their assembly, as previously reported (Nelson et al., 2020; Meziti et al., 2021).



All MAGs studied here were >95% complete with <5% contamination, a conventional criterion marking their high quality. Given the extreme importance of the 16S gene in microbial taxonomy and ecology, it seems unacceptable that MAGs can be labeled as such and at the same time contain low-quality information about this single most important gene that links the re-constructed genomes to the huge body of 16S-based microbiota studies conducted worldwide.

Furthermore, the acceptance of poor 16S rDNA quality in MAGs currently excludes a majority in the microbial research community that does not have the economic or computational resources to perform large-scale shotgun sequencing.

AUTHOR CONTRIBUTIONS

KR and PH conceived the idea. PH wrote the manuscript with an equal input from all authors. All authors discussed and interpreted the findings. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by Norway Research Council, a Norwegian government agency funding research and innovation, through R&D project grant nos. 283783, 248792, and 301364.

REFERENCES

- Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., et al. (2019). A new genomic blueprint of the human gut microbiota. *Nature* 568, 499–504. doi: 10.1038/s41586-019-0965-1
- Hiseni, P., Rudi, K., Wilson, R. C., Hegge, F. T., and Snipen, L. (2021). HumGut: a comprehensive human gut prokaryotic genomes collection filtered by metagenome data. *Microbiome* 9:165. doi: 10.1186/s40168-021-01114-w
- Hugerth, L. W., Larsson, J., Alneberg, J., Lindh, M. V., Legrand, C., Pinhassi, J., et al. (2015). Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* 16,279. doi: 10.1186/s13059-015-0834-7
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9,5114. doi: 10.1038/s41467-018-07641-9
- Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi: 10.1099/ijs.0.059774-0
- Louca, S., Doebeli, M., and Parfrey, L. W. (2018). Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome* 6,41. doi: 10.1186/s40168-018-0420-9
- Meziti, A., Rodriguez-R, L. M., Hatt, J. K., Peña-Gonzalez, A., Levy, K., et al. (2021). The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample. *Appl. Environ. Microbiol.* 87, e02593–e02520. doi: 10.1128/AEM.02593-20
- Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S., and Kyrpides, N. C. (2019). New insights from uncultivated genomes of the global human gut microbiome. *Nature* 568, 505–510. doi: 10.1038/s41586-019-1058-x
- Nelson, W. C., Tully, B. J., and Mobberley, J. M. (2020). Biases in genome reconstruction from metagenomic data. *PeerJ* 8,e10119. doi: 10.7717/peerj.10119
- Paradis, E., and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633
- Pasoli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., et al. (2019). Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 176, 649–662.e620. doi: 10.1016/j.cell.2019.01.001
- Perisin, M., Vetter, M., Gilbert, J. A., and Bergelson, J. (2016). 16Stimator: statistical estimation of ribosomal gene copy numbers from draft genome assemblies. *ISME J.* 10, 1020–1024. doi: 10.1038/ismej.2015.161
- Yuan, C., Lei, J., Cole, J., and Sun, Y. (2015). Reconstructing 16S rRNA genes in metagenomic data. *Bioinformatics* 31, i35–i43. doi: 10.1093/bioinformatics/btv231

Conflict of Interest: PH and KF were employed by company Genetic Analysis AS.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Hiseni, Snipen, Wilson, Furu and Rudi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.