Title

From soil to sequence: filling the critical gap in genome-resolved metagenomics is essential to the future of soil microbial ecology

Permalink

https://escholarship.org/uc/item/8tw2k368

Journal

Environmental Microbiome, 19(1)

ISSN

2524-6372

Authors

Anthony, Winston E
Allison, Steven D
Broderick, Caitlin M
et al.

Publication Date

2024

DOI

10.1186/s40793-024-00599-w

Peer reviewed

COMMENT

Open Access

From soil to sequence: filling the critical gap in genome-resolved metagenomics is essential to the future of soil microbial ecology

Winston E. Anthony[1*], Steven D. Allison[2,3], Caitlin M. Broderick[4], Luciana Chavez Rodriguez[2], Alicia Clum[5], Hugh Cross[6], Emiley Eloe-Fadrosh[5], Sarah Evans[4], Dawson Fairbanks[7,8], Rachel Gallery[8], Júlia Brandão Gontijo[9], Jennifer Jones[4], Jason McDermott[1], Jennifer Pett-Ridge[10,11], Sydne Record[12], Jorge Luiz Mazza Rodrigues[5,9], William Rodriguez-Reillo[13], Katherine L. Shek[14], Tina Takacs-Vesbach[15] and Jeffrey L. Blanchard[16*]

Abstract

Soil microbiomes are heterogeneous, complex microbial communities. Metagenomic analysis is generating vast amounts of data, creating immense challenges in sequence assembly and analysis. Although advances in technology have resulted in the ability to easily collect large amounts of sequence data, soil samples containing thousands of unique taxa are often poorly characterized. These challenges reduce the usefulness of genome-resolved metagenomic (GRM) analysis seen in other fields of microbiology, such as the creation of high quality metagenomic assembled genomes and the adoption of genome scale modeling approaches. The absence of these resources restricts the scale of future research, limiting hypothesis generation and the predictive modeling of microbial communities. Creating publicly available databases of soil MAGs, similar to databases produced for other microbiomes, has the potential to transform scientific insights about soil microbiomes without requiring the computational resources and domain expertise for assembly and binning.

Keywords Soil Microbiome, Hybrid Assembly, Genome Resolved Metagenomics, Microbiome Assembled Genomes, FAIR Data Principles

*Correspondence:
Winston E. Anthony
winston.anthony@pnnl.gov
Jeffrey L. Blanchard
jlb@umass.edu
[1]Pacific Northwest National Laboratory, Richland, WA 99354, USA
[2]University of California Irvine, Irvine, CA, USA
[3]Department of Earth System Science, University of California, Irvine, CA, USA
[4]W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI, USA
[5]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[6]National Ecological Observatory Network - Battelle, Boulder, CO, USA
[7]University of California Riverside, Riverside, CA, USA
[8]The University of Arizona, Tucson, AZ, USA
[9]University of California Davis, Davis, CA, USA
[10]Lawrence Livermore National Laboratory, Livermore, CA, USA
[11]Life & Environmental Sciences Department, University of California Merced, Merced, CA 95343, USA
[12]University of Maine, Orono, ME, USA
[13]Harvard Medical School, Boston, MA, USA
[14]University of New Hampshire, Durham, NH, USA
[15]University of New Mexico, Albuquerque, NM, USA
[16]University of Massachusetts Amherst, Amherst, MA, USA

Introduction

Soil microbial communities are incredibly diverse, and they provide crucial services such as supporting plant life, regulating human health [1], and driving global carbon cycling. Soil harbors an immense pool of carbon in the form of roots and decomposed organic matter that exceeds the amount of carbon aboveground in terrestrial plants and the atmosphere combined [2], and in the face of global change it is important to understand how soil microbial communities regulate carbon cycling and other functions. However, the soil microbiome remains largely undescribed; most taxa do not have sequenced genomes, isolation and cultivation of the entire diversity of soil bacteria and archaea is not currently feasible, and, with some exceptions, the viral and eukaryotic components are often not considered [3]. Identifying novel microbial taxa and understanding their influence on microbial diversity and ecosystem-level processes could shed light on methods for mitigating the effects of climate change [4–6].

Shotgun metagenomics has revolutionized our ability to examine complex patterns of functional and taxonomic diversity in soil. Using short read-based sequencing technologies (typically 75–250 base pairs), quality-filtered DNA sequences are used as input to search reference databases for gene function and taxonomy assignments [7]. Taxonomic assignments usually rely on the presence of marker regions, while mapping reads to databases of annotated genes or pathways provides an estimate of metabolic pathway coverage at the community level. Therefore, the estimation of important gene functions and community composition are reliant on separate, independently created and managed databases. A major limitation of this approach is that a scarcity of representative soil bacterial genomes and a lack of robust knowledge of their metabolic capabilities makes it difficult to link community structure with metabolic function.

The Human Microbiome Project (HMP) overcame this limitation on analyzing complex microbiome samples through the compilation of reference genomes and databases, which was a large upfront investment with big impact [8]. This comprehensive catalog of microbiome reference genomes results in the mapping of most human-associated metagenomic reads directly to the HMP genomes stored in Integrated Microbial Genomes & Microbiomes (IMG) from Earth's Microbiomes (GEM) and NCBI databases [9] (Fig. 1), enabling

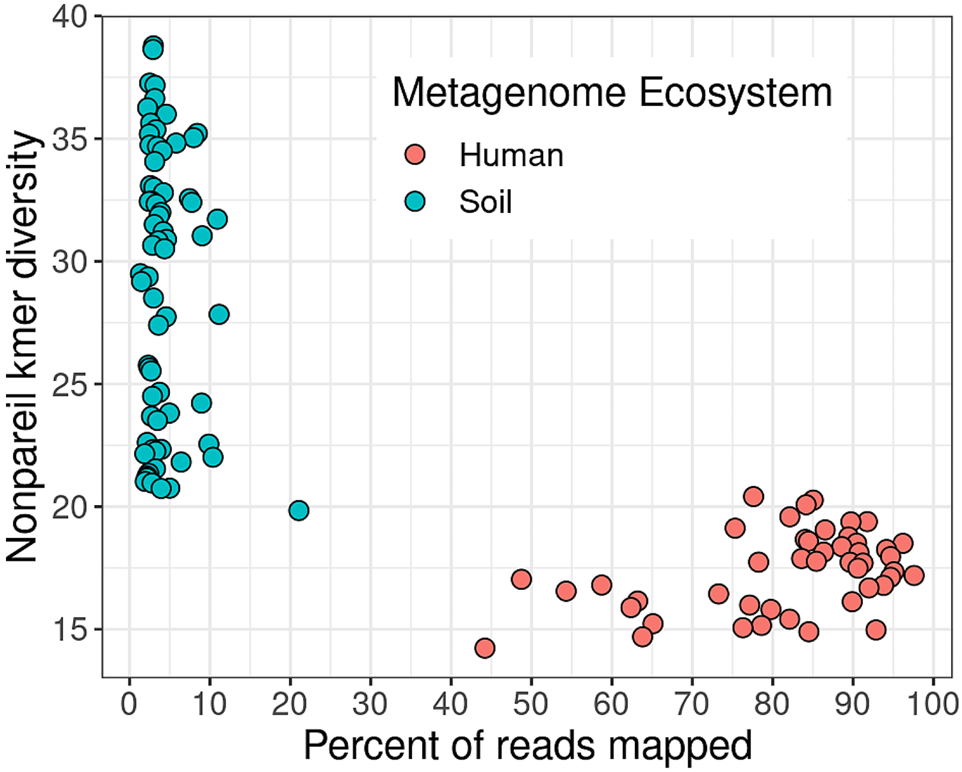

Fig. 1 Percent of soil and human metagenomic reads mapped to the GEM and NCBI RefSeq databases as a function of nonpareil kmer diversity. Nonpareil kmer diversity is a measure of genetic diversity within a metagenome. Human microbiomes are less diverse than soil (agricultural, forest and desert) metagenomes. From the efforts of the Human Microbiome Project there is a large collection of bacterial genome sequences in NCBI's RefSeq database and consequently a large proportion of reads from human metagenomes map to this database. Typically 50–90% of human metagenome reads map to the combined databases. In contrast very few soil metagenomic reads map to genomes in NCBI's RefSeq, and less than 5% map to the GEM catalog

the development of novel analysis methods and software tools for human microbiome analysis [10, 11].

Alongside the increasing availability of computational resources, improved read assembly into longer contiguous stretches (contigs) has increased the recovery of metagenome assembled genomes (MAGs) [12–14]. MAGs from an individual study can be a subset of the full catalog of larger public MAG databases and isolated genomes. By mapping reads to these MAGs, it is possible to link functional genes to specific organisms, providing a platform to bridge community structure with function [15] and enable finer-scale interrogations of emergent properties such as metabolite sharing and carbon cycling of environmental samples at both organism- and population levels. In this perspective we discuss the current state of GRM in soil microbiome research, cover computational and experimental advances propelling the discovery of microbial genomes from soil samples, and finally propose a roadmap towards an open and accessible database of genome-resolved soil microbiome sequences.

Genome-resolved metagenomics for soil microbiomes: current state of the field

Largely, the GRM approach has enjoyed greater use within the human microbiome field and other environments, where lower diversity enables better MAG construction and a higher percentage of read mapping. High microbial diversity in soils makes it difficult to resolve complete genomes from metagenomic samples [16], requiring much deeper sequencing and high per-sample costs for retrieving soil MAGs (Fig. 1). However, novel binning (the process of separating metagenomic reads into organism-specific groups) strategies involving sophisticated *co-* [17] and mixed- [18] assembly and bin refinement [19] strategies have arisen to increase resolving power. The resulting improvement in coverage of individual genomes within the metagenome can increase the number and quality of genome bins. Publicly available and easily accessible MAGs enable reuse by other researchers without requiring the computational resources and domain expertise for assembly and binning.

The recovery of high-quality draft genomes of previously uncharacterized viral and eukaryotic MAGs, though initially understudied, is increasingly an area of interest in soil systems. A recent study described novel giant viruses derived from soil samples from the Harvard Forest LTER site using Fluorescence-activated Cell Sorting (FACS) sorting [20]. Another study curated a database including 726,108 de-replicated viral contigs combining metagenome assembled contigs using prairie soil and public virus databases [3]. A terabase-scale combined assembly from Luquillo Experimental Forest, Puerto Rico, revealed tens of thousands of viruses and tens of partial eukaryotes [21]. These efforts contribute to expanding taxonomic databases such as International Committee on Taxonomy of Viruses (ICTV) [22] and National Center for Biotechnology Information (NCBI) Taxonomy Database [23]. IMG has implemented workflows to analyze viruses and eukaryotes as part of its routine processing. Efforts such as these are critical for building large, increasingly complete databases of soil microbial genomes, expanding our understanding of soil taxa, and facilitating efforts to link genes to specific microorganisms.

Efforts are underway to create environment-specific MAG databases and resources, albeit for less diverse ecosystems. The TARA oceans dataset was leveraged to generate thousands of MAGs from marine environments, while the Genome Resolved Open Watersheds database (GROWdb) focuses on microbes from rivers and streams [24–26]. Resources such as MGnify [27] and organizations such as the National Microbiome Data Collaborative (NMDC) [28] link to or host some of these catalogs, but there are still many environments such as soils that are significantly underrepresented and under sampled. There is a pressing need for high quality soil GRM databases, similar to the Human Microbiome Project that spans soils across space and time building upon collections of cultured soil representatives [29]. A recent step forward involved the reprocessing of soil-related metagenomes into a standalone set of data products including over 40,000 MAGs [30], producing the SMAG catalogue for future use.

The future of GRM for soil microbiomes

The complexity of soil microbiomes poses several technical hurdles to genome-resolved analyses. Soil is abundant with "relic DNA" [31, 32], defined as extracellular DNA from dead bacterial and fungal cells, which can hide temporal differences in sample from the same community. This can affect estimates of diversity [33]. Metatranscriptomics, which uses the quickly degraded RNA molecule, can be used to assess the functioning members of a microbial community [34]. Though there are DNA intercalation agents which can bind to and remove cell free DNA [31], relic DNA removal and quantification of its effect on analysis is still a developing area of soil microbiome research [32, 33].

Gaps in coverage and repetitive elements in a genome can cause fragmentation in assemblies, leading to incomplete and contaminated sequence bins. The use of long-read sequence platforms such as Oxford Nanopore and Pacific Biosciences, which can span typical microbial repeat lengths of 5–7 Kbp [8], can improve assembly and binning while reducing contamination. These platforms are more error-prone, but are becoming less expensive and more mainstream [35]. Hybrid sequence analysis approaches utilizing long and short reads are now capable of producing higher quality, more contiguous assemblies than either technique alone at lower depth [36]. Tools such as SPAdes (hybridSPAdes) [37] and Unicycler [38] both utilize short reads to produce an initial assembly graph and then close or bridge gaps with contigs assembled from long reads. Multiple studies have attempted to compare the abilities of short, long, and hybrid sequencing approaches for de novo MAG catalog creation [39–43], reporting differences in GC distributions of recovered MAGs, and in the number of detected genes [40]. Most recently, Eisenhofer et al. report that while short reads capture more diversity in recovered MAGs due to higher sequence depth, long read and hybrid assembly strategies result in better assembly statistics. They end their analysis by suggesting that the optimal sequencing/assembly strategy is highly study-specific and will change based on whether MAG quantity or quality is deemed more valuable [43].

Algorithmic advances are improving results for the same input data [44]. For example, machine learning approaches that increase protein function identification are being applied to sequencing data [45, 46] and assembly [47]. Hi-C (capture chromatin conformation) and similar technologies allow for binning based on physical proximity rather than tetranucleotide frequency or sequencing coverage, greatly increasing the ability to identify sequences from the same genome and is used to identify bacteria-phage associations [48, 49]. Finally, there are tools purpose-built to identify contamination and chimerism in prokaryotic genomes, such as checkM [50], and GUNC [51].

An alternative set of approaches aim to reduce complexity through segregation and sorting. Methods such as FACS [52] and Stable-Isotope Probing (SIP) [53] segregate and reduce complex communities, lowering the sequence depth necessary to recover high quality genome bins. Novel viruses identified in samples collected from Harvard Forest LTER were derived from FACS-sorted samples, demonstrating the utility of artificially simplifying complex environments for viral genome enrichment from soil samples [20]. SIP-based methods are traditionally very labor intensive, but recent work has improved throughput [53].

Genome-resolved analyses improve our ability to explore complex soil communities where species' physiology is unknown and culture-based methods are infeasible. GRM can achieve increased taxonomic resolution for the large diversity of environmental microbes but require robust databases of cumulative genomic knowledge. The fungal component of soil microbiomes is understudied, but recent advances in long read sequencing are opening the field, leading to new insight into fungal diversity and evolution [54]. Efforts to include fungal and eukaryotic microbes in microbiome catalogs are occurring in other environments [55, 56], but soil-specific fungal MAG catalogs are needed. As stated previously, community-driven projects such as TARA, GROWdb, MGnify, NMDC, SMAG, FUNGIDB [57], etc. are excellent initial efforts, and we propose the following goals which should be met or exceeded to build on this foundation:

- Curation: A future soil MAG database needs to contain sampling throughout the entire spectrum of the soil medium. An excellent start is seen in the Joint Genome Institute's GOLD organism ecosystem classifications, however currently 15,154/23,473 (64.6%) of soil organisms are not classified into a specific ecosystem subtype. A future database requires full FAIR metadata schema compliance [58] and version control for the entirety of its data processing. The creation of a new soils-specific GO FAIR implementation network (https://www.go-fair.org) could work to generate microbiome-specific FAIR practices and tools. We currently recommend following the latest guidance, Minimum Information about any (X) Sequence (MIxS) version 6.2.0, provided by the Genome Standards Consortium (GSC) [59].
- Scale: Initial large scale efforts [8, 60, 61] to survey the human gut microbiome rewarded up-front investment. To recreate those successes in soil microbial ecology, a comprehensive, uniform survey of many different soil types and environments is required, such as the new MONET initiative by the Department of Energy's Environmental Molecular Sciences Laboratory (EMSL) [62].
- Integration: Ease of use is frequently a barrier to community adoption of new methods and datasets. A future soil MAG database should feature easy integration and data forwarding into KBASE [63], GALAXY [64], The NMDC [28], and other data pipelines and analysis centers. Furthermore, greater acceptance and use of standardized community tools will increase analysis re-producibility. NMDC EDGE (https://microbiomedata.org/workflows/) is a user-friendly web interface community members can use to process their own data in a standardized fashion using community-agreed upon metrics. Tutorials are provided in several languages.

Adoption of these principles in the creation of a MAG database at this scale would require substantial upfront investment. However, in studies where GRM are infeasible due to practical constraints such as cost and sample size, an added benefit of a large MAG database is to make existing and future taxa-based (amplicon) datasets more accurate. Through collaborative efforts to increase the number of high-quality reference MAGs, we can advance our understanding of the biodiversity and ecology of one of Earth's most complex environments.

Acknowledgements

We would like to acknowledge and thank the efforts of Margaret O'Brien at EDI. This work would not have been possible if not for the support of the entire LTER EMERGENT working group, and the National Center for Ecological Analysis & Synthesis in Santa Barbara, California.

Author contributions

Conceptualization: W.E.A, A.C, K.S., C.B, J.B, L.C.R, J.B.G, D.E.F, J.P.R, H.C. Funding Acquisition: J.B.Methodology: W.E.A. Project Administration: W.E.A. Resources: W.E.A. Visualization: J.B.G, K.S, L.C.R, J.B. Writing - Original Draft: W.E.A, A.C, K.S., C.B, J.B, L.C.R, D.E.F. Writing - Review & Editing: All authors reviewed the manuscript.

Funding

This research was supported by the U.S. Department of Energy (DOE) Office of Biological and Environmental Research (BER) as part of BER's Genomic Science Program (GSP) and is a contribution of the Pacific Northwest National Laboratory (PNNL) Secure Biosystems Design Science Focus Area "Persistence Control of Engineered Functions in Complex Soil Microbiomes". PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830. This material is based upon work supported by the National Science Foundation under grant DEB#1545288, 10/1/2015-9/30/19 and DEB#1929393, 09/01/2019-08/31/2024. Work at Lawrence Livermore National Laboratory (LLNL) was conducted under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, SCW1632 and contract DE-AC52-07NA27344. This work is supported by the U.S. Department of Energy Office of Science, through the Genomic Science Program in the Office of Biological and Environmental Research (BER). The national laboratory partners are operated under contract numbers DE-AC02-05CH11231 (LBNL), 89233218CNA000001 (LANL), and DE-AC05-76RL01830 (PNNL).

Data availability

Figure 1 is derived from supplemental materials in Nayfach et al. 2021 [9]. Code to generate the figure is available in the *EMERGENT github repository* https://github.com/lter/lterwg-emergent/tree/master/perspective.

Declarations

Competing interests

The authors declare no competing interests.

Received: 12 May 2024 / Accepted: 22 July 2024
Published online: 02 August 2024

References

1. Banerjee S, van der Heijden MGA. Soil microbiomes and one health. Nat Rev Microbiol. 2023;21:6–20.
2. Scharlemann JP, Tanner EV, Hiederer R, Kapos V. Global soil carbon: understanding and managing the largest terrestrial carbon pool. Carbon Manag. 2014;5:81–91.
3. Wu R, et al. Moisture modulates soil reservoirs of active DNA and RNA viruses. Commun Biol. 2021;4:992.
4. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol. 2017;15:579–90.
5. Wagg C, Schlaeppi K, Banerjee S, Kuramae EE, van der Heijden MG. A. Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. Nat Commun. 2019;10:4841.
6. Bastida F, et al. Soil microbial diversity-biomass relationships are driven by soil carbon content across global biomes. ISME J. 2021;15:2081–91.
7. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017;35:833–44.
8. Huttenhower C, et al. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486:207–14.
9. Nayfach S, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021;39:499–509.
10. Segata N, et al. Computational meta'omics for microbial community studies. Mol Syst Biol. 2013;9:666.
11. Beghini F, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife. 2021;10:e65088.
12. Tyson GW, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 2004;428:37–43.
13. Setubal JC. Metagenome-assembled genomes: concepts, analogies, and challenges. Biophys Rev. 2021;13:905–9.
14. Sangwan N, Xia F, Gilbert JA. Recovering complete and draft population genomes from metagenome datasets. Microbiome. 2016;4:8.
15. Singleton CM et al. Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nat. Commun* 12, 2009 (2021).
16. Nelson WC, et al. Terabase Metagenome sequencing of Grassland Soil Microbiomes. Microbiol Resour Announc. 2020;9. https://doi.org/10.1128/mra.00718.
17. Hofmeyr S, et al. Terabase-scale metagenome coassembly with MetaHipMer. Sci Rep. 2020;10:10689.
18. Delgado LF, Andersson AF. Evaluating metagenomic assembly approaches for biome-specific gene catalogues. Microbiome. 2022;10:1–11.
19. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6:158.
20. Schulz F, et al. Hidden diversity of soil giant viruses. Nat Commun. 2018;9:4881.
21. Riley R et al. Terabase-Scale Coassembly of a Tropical Soil Microbiome. Microbiol Spectr e00200–23 (2023).
22. Davison AJ. Journal of General Virology – introduction to 'ICTV Virus Taxonomy profiles'. J Gen Virol. 2017;98:1–1.
23. Sayers EW, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022;50:D20–6.
24. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data. 2018;5:1–8.
25. Borton MA et al. A functional microbiome catalog crowdsourced from North American rivers. *bioRxiv* 2023–07 (2023).
26. Sunagawa S, et al. Tara Oceans: towards global ocean ecosystems biology. Nat Rev Microbiol. 2020;18:428–45.
27. Gurbich TA, et al. MGnify genomes: a resource for Biome-specific Microbial Genome catalogues. Comput Resour Mol Biol. 2023;435:168016.
28. Eloe-Fadrosh EA, et al. The National Microbiome Data Collaborative Data Portal: an integrated multi-omics microbiome data resource. Nucleic Acids Res. 2022;50:D828–36.
29. Choi J, et al. Strategies to improve reference databases for soil microbiomes. ISME J. 2017;11:829–34.
30. Ma B, et al. A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources. Nat Commun. 2023;14:7318.
31. Carini P, et al. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nat Microbiol. 2016;2:1–6.
32. Lennon JT, Muscarella ME, Placella SA, Lehmkuhl BK. How, when, and where Relic DNA affects Microbial Diversity. mBio. 2018;9. https://doi.org/10.1128/mbio.00637-18.
33. Carini P, et al. Effects of spatial variability and Relic DNA removal on the detection of temporal dynamics in Soil Microbial communities. mBio. 2020;11. https://doi.org/10.1128/mbio.02776-19.
34. Zaikova E et al. Antarctic Relic Microbial Mat Community revealed by Metagenomics and Metatranscriptomics. Front Ecol Evol 7, (2019).
35. Marx V. Method of the year: long-read sequencing. Nat Methods. 2023;20:6–11.
36. Chen Z, Erickson DL, Meng J. Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. BMC Genomics. 2020;21:631.
37. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics. 2016;32:1009–15.
38. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput Biol. 2017;13:e1005595.

39. Tao Y et al. Improved Assembly of Metagenome-assembled genomes and viruses in tibetan saline Lake Sediment by HiFi Metagenomic sequencing. Microbiol Spectr 11, e03328–22.
40. Orellana LH, Krüger K, Sidhu C, Amann R. Comparing genomes recovered from time-series metagenomes using long- and short-read sequencing technologies. Microbiome. 2023;11:105.
41. Gehrig JL, et al. Finding the right fit: evaluation of short-read and long-read sequencing approaches to maximize the utility of clinical microbiome data. Microb Genomics. 2022;8:000794.
42. Zhang Z, Yang C, Veldsman WP, Fang X, Zhang L. Benchmarking genome assembly methods on metagenomic sequencing data. Brief Bioinform. 2023;24:bbad087.
43. Eisenhofer R, et al. A comparison of short-read, HiFi long-read, and hybrid strategies for genome-resolved metagenomics. Microbiol Spectr. 2024;12:e03590–23.
44. Zhang H, Jain C, Aluru S. A comprehensive evaluation of long read error correction methods. BMC Genomics. 2020;21:889.
45. Bonetta R, Valentino G. Machine learning techniques for protein function prediction. Proteins Struct Funct Bioinforma. 2020;88:397–413.
46. Sanderson T, Bileschi ML, Belanger D, Colwell LJ. ProteInfer: deep networks for protein functional inference. *Biorxiv* 2021–09 (2021).
47. de Padovani K, et al. Machine learning meets genome assembly. Brief Bioinform. 2019;20:2116–29.
48. Wu R, et al. Hi-C metagenome sequencing reveals soil phage–host interactions. Nat Commun. 2023;14:7666.
49. Belton J-M, et al. Hi–C: a comprehensive technique to capture the conformation of genomes. Methods. 2012;58:268–76.
50. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
51. Orakov A, et al. GUNC: detection of chimerism and contamination in prokaryotic genomes. Genome Biol. 2021;22:178.
52. Bonner WA, Hulett HR, Sweet RG, Herzenberg LA. Fluorescence activated cell sorting. Rev Sci Instrum. 2003;43:404–9.
53. Nuccio EE, et al. HT-SIP: a semi-automated stable isotope probing pipeline identifies cross-kingdom interactions in the hyphosphere of arbuscular mycorrhizal fungi. Microbiome. 2022;10:199.
54. Priest SJ, Yadav V, Heitman. J. Advances in understanding the evolution of fungal genome architecture. *F1000Research* 9, F1000 Faculty Rev-776 (2020).
55. Peng X, et al. Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. Nat Microbiol. 2021;6:499–511.
56. Singh NK, et al. Characterization of metagenome-assembled genomes from the International Space Station. Microbiome. 2023;11:125.
57. Stajich JE, et al. FungiDB: an integrated functional genomics database for fungi. Nucleic Acids Res. 2012;40:D675–81.
58. Wilkinson MD, et al. The FAIR Guiding principles for scientific data management and stewardship. Sci Data. 2016;3:160018.
59. Yilmaz P, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol. 2011;29:415–20.
60. Zeng S, et al. A compendium of 32,277 metagenome-assembled genomes and over 80 million genes from the early-life human gut microbiome. Nat Commun. 2022;13:5139.
61. Qin J, et al. A human gut microbial gene catalog established by metagenomic sequencing. Nature. 2010;464:59–65.
62. Molecular Observation Network (MONet). | Environmental Molecular Sciences Laboratory. https://www.emsl.pnnl.gov/monet.
63. Arkin AP, et al. KBase: the United States department of energy systems biology knowledgebase. Nat Biotechnol. 2018;36:566–9.
64. The Galaxy Community. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res. 2022;50:W345–51.

Publisher's Note