

RESEARCH

Open Access



Automated environmental metagenomics using Oxford nanopore sequencing

Harry T. Child¹, Lucy Wierzbicki¹, Gabrielle R. Joslin¹, Katherine Roper², Qiellor Haxhiraj² and Richard K. Tennant^{1*}

Abstract

Background Long-read sequencing has revolutionised metagenomics through improved metagenome assembly, taxonomic classification and functional characterisation. Automation can enhance the throughput, reproducibility, and accuracy of library preparation. However, the validation of automated library preparation protocols remains undetermined for metagenomic workflows, which are particularly sensitive to methodological perturbation. Here, we compare long-read metagenomic sequencing of environmental samples through parallel manual and automated protocols.

Results Although automated library preparation led to minor reduction in read and contig lengths, taxonomic classification rate and alpha diversity was slightly higher than manual libraries, including the detection of more rare taxa. Despite this, no significant difference in microbial community structure was identified between manual and automated libraries.

Conclusions Despite minor differences in sequencing and classification metrics, automated and manual library preparation resulted in comparable characterization of environmental community metagenomes. These findings demonstrate the suitability of automation for high-throughput long-read metagenomics, with broad applicability to automated long-read sequencing for improved efficiency and reproducibility.

Keywords Automation, Long-read sequencing, Metagenomics, Oxford nanopore, Library preparation, Soil

Background

Long-read sequencing has transformed our understanding of microbiomes through improved genome assembly, functional characterisation and taxonomic classification accuracy and precision [1–3], leading to its rapid expansion in metagenomic research [4]. Third generation sequencing methods from both Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) have

enabled the assembly of complete bacterial genomes from environmental and host-associated microbiomes [1, 5–7], allowing confident functional classification of unculturable microorganisms [8, 9]. Furthermore, utilisation of full-length 16 S amplicon sequencing and longer whole genome metagenomic reads provides more information for higher resolution of taxonomic classification [2, 10–12]. The enhanced capacity for multiplexing samples on Oxford Nanopore Technologies (ONT) platforms, as well as increases in potential yield and reduced costs of ONT sequencing [4], has improved the potential throughput of long read metagenomics. However, multiplexed library preparation protocols involve many pipetting steps, requiring considerable hands-on time and

*Correspondence:

Richard K. Tennant

R.K.Tennant@exeter.ac.uk

¹Geography, Faculty of Environment, Science and Economy, Amory Building, Rennes Drive, Exeter, Devon EX4 4RJ, UK

²Agilent Technologies LDA UK Limited, 5500 Lakeside, Cheadle Royal Business Park, Stockport, Cheshire SK8 3GR, UK



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

introducing potential for human error and inter-sample variation.

Automation using liquid handling robotics therefore has the potential to enhance the throughput, reproducibility, and accuracy of sequencing library preparation [13, 14]. However, validation of liquid handling automation for ONT protocols remains limited in the literature, with studies limited to high throughput amplicon sequencing of SARS-CoV-2 [15]. Along with clinical applications, validation of sample preparation processes is particularly important in metagenomic workflows due to the sensitivity of these analyses to perturbations from methodological bias [16, 17], which can impact the interpretation of study results [18, 19].

Here, we compared long-read metagenomic sequencing of environmental samples using either manual or automated ONT library preparation. We utilised the Bravo Automated Liquid Handling Platform (Agilent Technologies, UK), which has a 96-channel pipetting head for simultaneous execution of liquid handling steps across a 96-well plate. ONT sequencing libraries were prepared in parallel manually and on the Bravo using 24 DNA samples, extracted from soils with a range of habitat and geochemical traits. Analysis of metagenomic data revealed that while there were differences in read length, classification rate and alpha diversity between manual and automated libraries, there was minimal impact on the observed microbial community composition. Considering the benefits of reduced hands-on time, reproducibility and reliability, automated library preparation using the Bravo should be considered for increasing throughput of long-read sequencing.

Results and discussion

Sequencing read metrics were compared between automated and manual library preparations (Fig. 1; Supplementary Table 1). No significant difference in sequencing depth or read quality scores was identified between paired libraries (Fig. 1a-b; Supplementary Table 1). However, read length was found to be significantly longer from manually prepared libraries, with a mean difference in average length and N50 of 756 bp and 785 bp, respectively (Fig. 1c-d; Supplementary Table 1). This resulted in more contiguous assemblies from manually prepared libraries (Fig. 1e; Supplementary Table 1). However, when reads were taxonomically assigned using *Kraken2*, a small but significantly higher percentage of reads was classified from automated libraries (Fig. 1f; Supplementary Table 1), with a mean difference in classification rate of only 0.5% (excluding the outlying sample from the pasture soil).

Differences in read length and resulting assembly contiguity are likely caused by variation in bead purification steps between manual and automated protocols. While

shaking to elute DNA from magnetic beads was carried out at 37 °C in the manual protocol to improve elution of long fragments, as recommended in the ONT protocol, simultaneous temperature control and shaking was not possible on the Bravo. This may have caused reduced efficiency of long DNA fragment elution for the automated libraries, potentially reducing the average fragment length of automated libraries. However, the length of DNA was not analysed prior to pooling to confirm this. Meanwhile, the taxonomic classification rate may have been slightly improved in automated libraries through increased efficacy of DNA purification, leading to reduction in PCR artefacts.

Metagenome-assembled genomes (MAGs) were generated using assemblies from each sample and all medium- and high-quality MAGs were selected for further analysis. Overall, MAG recovery was poor, with 65% of libraries resulting in 0–1 medium-quality MAGs (Supplementary Table 1), likely due to insufficient sequencing depth. Although completeness and contamination across all MAGs was not significantly different between library preparation methods (Supplementary Fig. 1a-b), significantly more medium quality MAGs per sample were generated from manually prepared libraries (Supplementary Fig. 1c). This difference is mostly driven by the heath soil, which did not result in any medium quality MAGs from automated library preparation (Supplementary Table 1; Supplementary Fig. 1d). This soil type displayed the longest read and assembly lengths of any soil type in manually prepared libraries (Supplementary Table 1), likely facilitating improved MAG generation. Meanwhile, medium quality MAG assembly was improved in automated libraries from the woodland soil (Supplementary Table 1; Supplementary Fig. 1c-d), indicating soil type-driven variability in assembly from different library preparation methods. These results indicate some improvement in MAG generation using manually prepared libraries. However, considering long-read MAG assembly from complex metagenomic samples benefits from considerably higher sequencing depth (>20 Gbp per sample) [1, 20, 21], high-throughput metagenomic library preparation using automation is likely to be more suited to ecological analysis through taxonomic and functional classification.

Ecological analyses were performed on the results of taxonomic classification at a Family level (Fig. 2). A significant increase in alpha diversity, measured as both Shannon-Weaver index and family richness, was observed in libraries prepared on the Bravo (Fig. 2a-b), which was mostly the result of the presence of rare taxa (Fig. 2d). Detection of rare microorganisms in complex samples is an important objective of many metagenomic studies, due to their importance to ecosystem functions and community dynamics [22, 23], for which the

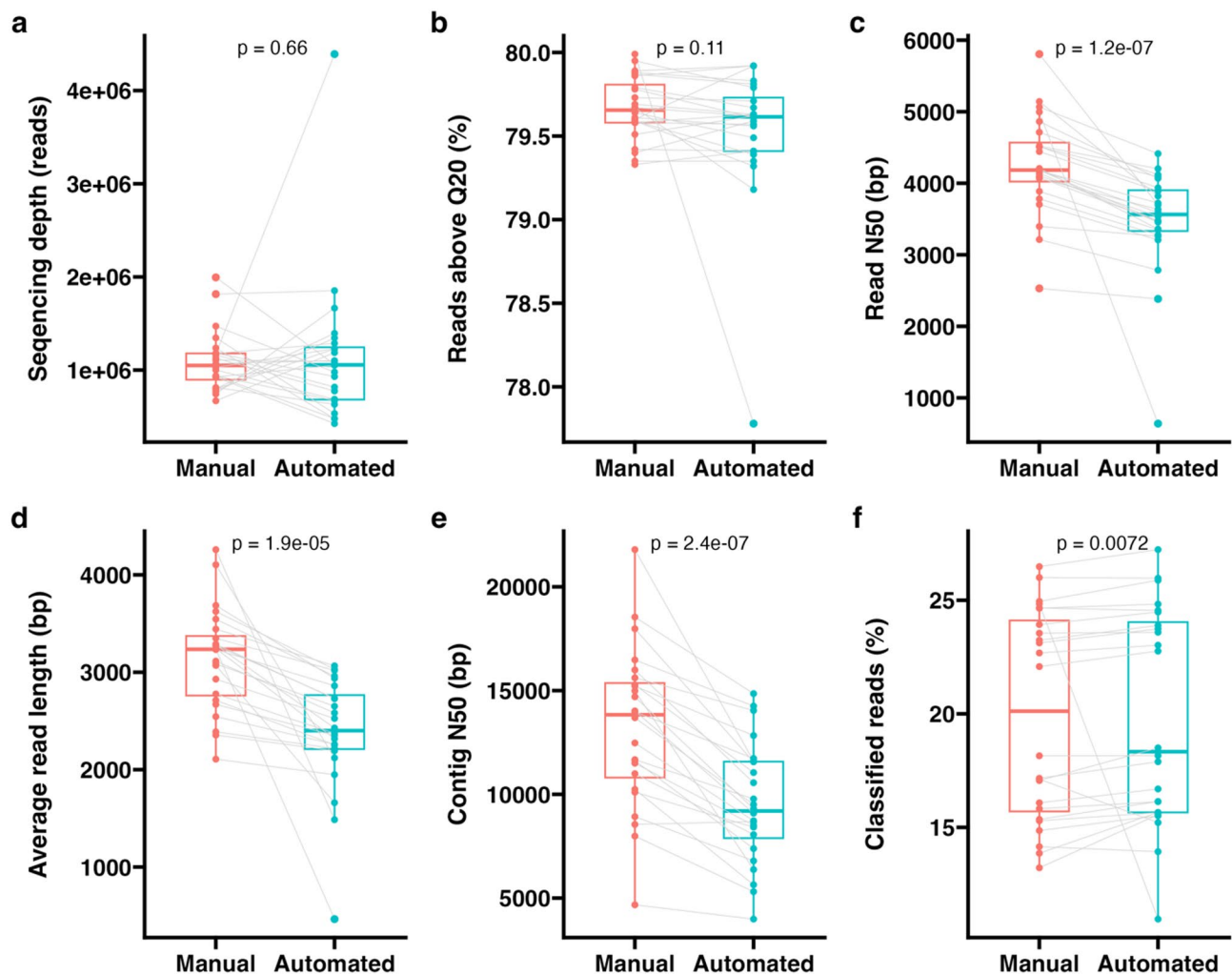


Fig. 1 Sequencing read metrics. Boxplots comparing (a) sequencing read depth, b read length N50, c mean read length, d percentage of reads above a quality score of Q20, e N50 of assembled contigs and (f) percentage of reads classified by Kraken2 from manual or automated library preparation. Grey lines indicate paired samples prepared in parallel and results of Wilcoxon signed-rank tests are displayed

increased diversity of automated libraries observed here could provide a benefit.

Variation in microbial community structure was investigated through calculation of Bray-Curtis distances with rarefaction (Fig. 2c). Unsurprisingly, soil type was found to explain the vast majority of variation in community composition between the samples (PERMANOVA, $R^2 = 0.92$, $p < 0.001$), while library preparation method or the interaction between these variables showed no significant effect (Fig. 2c). To support this, analysis within each soil type found no significant effect of library preparation method on microbial community composition at any taxonomic rank (PERMANOVA, $p > 0.05$). This indicates that minimal differences in microbial community composition were observed between manual and automated libraries, with no pattern to this variation within each soil type. Such consistency is crucial if the results from manual and automated library preparations are to

be compared, considering the importance of reproducibility for interpretation of metagenomic data within and between studies.

Demonstrating reproducibility is especially important for analysis of environmental samples, such as soil, that are particularly vulnerable to perturbation by methodological variation [16, 18]. The soil matrix exhibits high spatial heterogeneity of microorganism distribution [16], as well as containing an abundance of inhibitors posing a challenge to molecular genetic analysis. Furthermore, microbial ecologists wish to characterise soil communities from a field to a continental scale [24, 25], while most soil nucleic acid extraction methods require comparably minuscule input quantities (250 μg –2 g). Considering these factors, and the statistical analysis required for deciphering differential abundance, sufficient sampling sizes and replication are crucial to uncover patterns in microbial community composition and function between

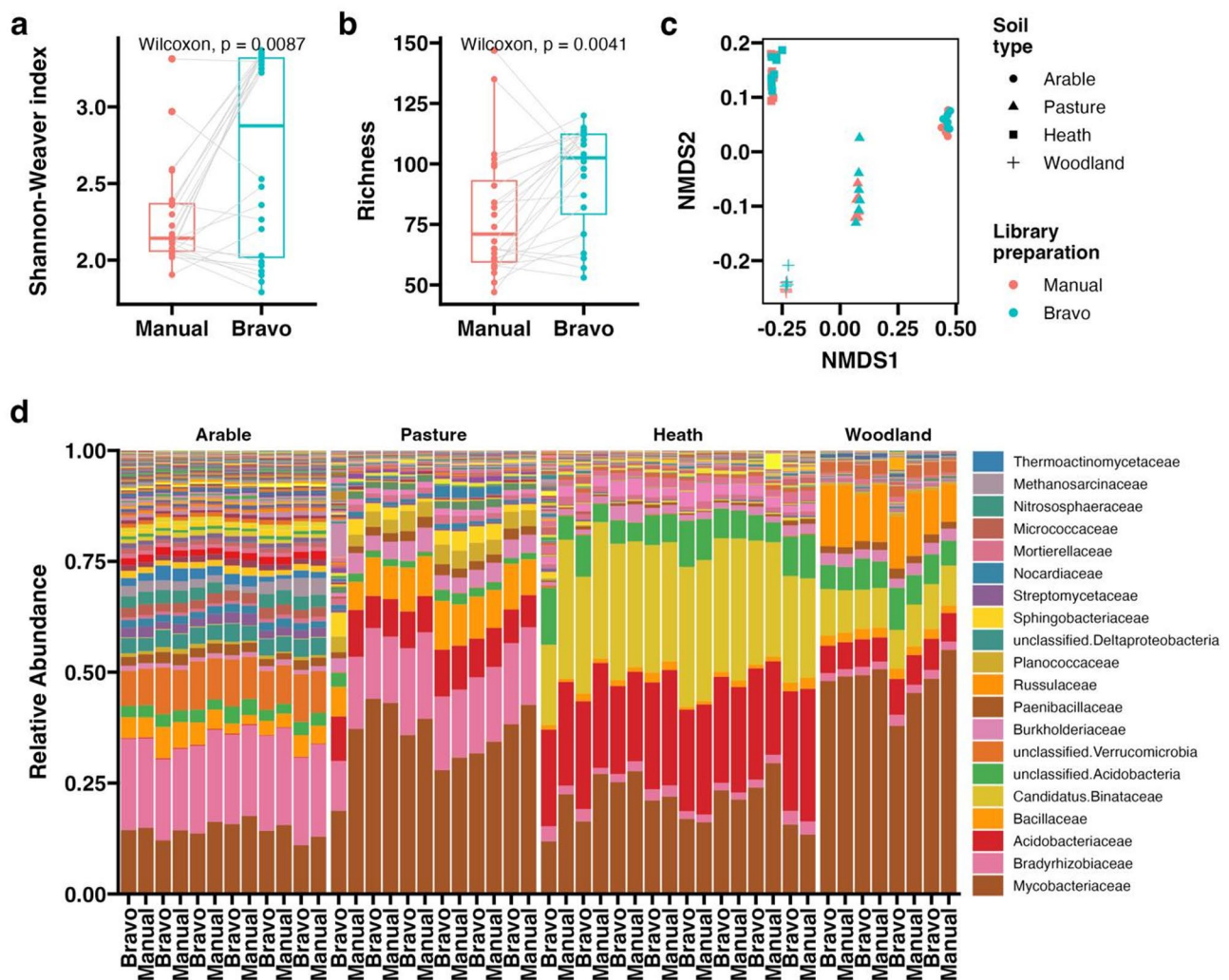


Fig. 2 Family level microbial community analysis. Boxplots comparing alpha diversity metrics calculated at the Family taxonomic rank, including (a) Shannon-Weaver index and (b) family richness, with grey lines indicating paired samples prepared in parallel and Wilcoxon signed-rank test results displayed. c Non-metric Multidimensional Scaling (nMDS) plot based on Bray-Curtis distances, showing variation between the observed microbial community structure of manual and automated libraries from the four soil types. d Stacked bar chart showing the relative abundance of microbial families across the four soil types. Legend shows colours corresponding to the top 20 families

sites and experimental treatments [16, 26, 27]. Automation has the potential to address these challenges of increased throughput and maintain reproducibility.

Conclusion

Despite the identification of minor differences in sequencing metrics and detection of rare taxa between manual and automated protocols, automated library preparation had minimal impact on the microbial community characterised from parallel metagenomic analysis of soil DNA samples. Considering the benefits of reduced hands-on time, reproducibility and reliability, automated library preparation should be considered suitable for improving throughput of ONT long-read sequencing.

Methods

Soil samples from four habitats were collected and characterised as previously described [28]. DNA was extracted using the DNeasy® PowerSoil® Pro Kit (Qiagen, UK), with 4–8 extractions from each soil type. DNA input into library preparations was normalised to 1 µg. Libraries were prepared using the Ligation Sequencing Kit (SQK-LSK114; ONT, UK) and PCR Barcoding Expansion 96 (EXP-PBC096; ONT, UK), with parallel preparations carried out manually, following manufacturer's protocol (Additional file 1), and automated on the Bravo (detailed in Additional file 2) on the same samples. Between 15 and 45 ng DNA was input into PCR barcoding reactions. Parallel preparations concluded with normalised pools of barcoded libraries, which were subsequently pooled

together and sequenced on the same R10.4.1 PromethION flowcell.

Sequencing yielded 125.74 Gb of data in 42.79 million reads with an N50 of 3.99 kb. Reads were basecalled and demultiplexed using *guppy* v7.1.4 and adapters were trimmed using *dorado* v0.6.0. Metagenome assembly was carried out using *metaFlye* [29], with reads input using the `--nano-hq` parameter. Contigs above 1 kb were binned with *SemiBin2* [30], using the `single_easy_bin` function (with the parameters `--sequencing-type=long_read` `--self-supervised`). The resulting MAGs were assessed for quality using *CheckM2* [31] and the filtered medium and high quality MAGs (genome completeness > 50% and contamination < 10%) were dereplicated to identify overlap between samples using *dRep* [32]. Taxonomic classification was carried out using *Kraken2* v 2.1.2 [33] against the NCBI nr database (downloaded on 09/03/24), using confidence score threshold of 0.05 to reduce the occurrence of false positives. Count tables were compiled using *MEGAN* Ultimate Edition v6.25.6 [34] and filtered to remove taxa occurring at an abundance of < 0.1% across all samples. Ecological statistics were calculated using the *vegan* v2.6-4 R package [35]. Bray-Curtis distances were calculated using the *avgdist* function with subsampling to the minimum classified read counts across samples.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11989-w>.

Additional file 1. Protocol checklist for ligation sequencing V14 with PCR barcoding (SQK-LSK114 with EXP-PBC001 or EXP-PBC096) on PromethION.

Additional file 2. ONT General Ligation Agilent Bravo Option B Automated User Guide

Additional file 3. Supplementary Figure 1. Analysis of medium-quality metagenome-assembled genomes. Supplementary Table 1. Table of read and assembly statistics for each sequenced sample.

Acknowledgements

Authors wish to thank Dr Sam Bridgewater and the Clinton Devon Estate for access to soils across different habitats and Dr Tomasz Dobrzycki for advice on Oxford Nanopore Technologies protocols.

Authors' contributions

HTC and RKT conceptualised the study. LW, GRJ, KR and QH optimised protocols and performed laboratory experimentation. HTC processed and analysed the sequencing data, prepared the figures and wrote the manuscript. RKT edited the manuscript. All authors read and approved the final manuscript.

Funding

This research was supported by Shell Research Ltd (CW648947-PT34767).

Data availability

The datasets generated and analysed during the current study are available in the NCBI Sequence Read Archive repository, [BioProject accession PRJNA1112790] (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1112790>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

KR and QH are employees of Agilent Technologies LDA UK Limited. All other authors declare that they have no competing interests.

Received: 15 July 2024 / Accepted: 8 August 2025

Published online: 26 September 2025

References

1. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, et al. Oxford nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods*. 2022;19(7):823–6.
2. Portik DM, Brown CT, Pierce-Ward NT. Evaluation of taxonomic classification and profiling methods for long-read shotgun metagenomic sequencing datasets. *BMC Bioinformatics*. 2022;23(1):541.
3. Chen Y, Nie F, Xie SQ, Zheng YF, Dai Q, Bray T, et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun*. 2021;12(1):60.
4. Agostinho DP, Fu Y, Menon VK, Metcalf GA, Treangen TJ, Sedlazeck FJ. Unveiling microbial diversity: harnessing long-read sequencing technology. *Nat Methods*. 2024;21:954–66.
5. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol*. 2020;38(6):701–7.
6. Sereika M, Mussig AJ, Jiang C, Knudsen KS, Jensen T, Petriglieri F et al. Recovery of highly contiguous genomes from complex terrestrial habitats reveals over 15,000 novel prokaryotic species and expands characterization of soil and sediment microbial communities [Internet]. *Bioinformatics*; 2024 [cited 2025 Jan 3]. Available from: <https://doi.org/10.1101/2024.12.19.629313>
7. Benoit G, Raguideau S, James R, Phillippy AM, Chikhi R, Quince C. High-quality metagenome assembly from long accurate reads with metamdbg. *Nat Biotechnol*. 2024;1–6.
8. Van Goethem MW, Osborn AR, Bowen BP, Andeer PF, Swenson TL, Clum A, et al. Long-read metagenomics of soil communities reveals phylum-specific secondary metabolite dynamics. *Commun Biol*. 2021;4(1):1–10.
9. Singleton CM, Petriglieri F, Kristensen JM, Kirkegaard RH, Michaelsen TY, Andersen MH, 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*. 2021;12(1):2009.
10. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*. 2019;10(1):5029.
11. Pearman WS, Freed NE, Silander OK. Testing the advantages and disadvantages of short- and long-read eukaryotic metagenomics using simulated reads. *BMC Bioinformatics*. 2020;21(1):220.
12. Lin X, Waring K, Ghezzi H, Tropini C, Tyson J, Ziels RM. High accuracy meets high throughput for near full-length 16S ribosomal RNA amplicon sequencing on the Nanopore platform. Benos P, editor. *PNAS Nexus*. 2024;3(10):pgae411.
13. Socea JN, Stone VN, Qian X, Gibbs PL, Levinson KJ. Implementing laboratory automation for next-generation sequencing: benefits and challenges for library preparation. *Front Public Health*. 2023;11.
14. Hess JF, Kohl TA, Kotrová M, Rönsch K, Paprotka T, Mohr V, et al. Library preparation for next generation sequencing: A review of automation strategies. *Biotechnol Adv*. 2020;41:107537.
15. Coope RJN, Matic N, Pandoh PK, Corbett RD, Smalish DE, Pleasance S, et al. Automated library construction and analysis for High-Throughput nanopore sequencing of SARS-CoV-2. *J Appl Lab Med*. 2022;7(5):1025–36.
16. Lombard N, Prestat E, van Elsas JD, Simonet P. Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol Ecol*. 2011;78(1):31–49.

17. Nearing JT, Comeau AM, Langille MGI. Identifying biases and their potential solutions in human Microbiome studies. *Microbiome*. 2021;9(1):113.
18. Changey F, Bland A, Pando A, Herrmann AM, Lerch TZ. Monitoring soil microbial communities using molecular tools: DNA extraction methods May offset long-term management effects. *Eur J Soil Sci*. 2021;72(2):1026–41.
19. Schloss PD. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in Microbiome research. *mBio*. 2018;9(3):10.
20. Benoit G, James R, Raguideau S, Alabone G, Goodall T, Chikhi R, et al. High-quality metagenome assembly from nanopore reads with NanoMDBG. *bioRxiv*; 2025.
21. Sereika M, Mussig AJ, Jiang C, Knudsen KS, Jensen TBN, Petriglieri F et al. Genome-resolved long-read sequencing expands known microbial diversity across terrestrial habitats. *Nat Microbiol*. 2025 Jul 24 [cited 2025 Jul 29]; Available from: <https://www.nature.com/articles/s41564-025-02062-z>
22. Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, et al. Conditionally rare taxa disproportionately contribute to Temporal changes in microbial diversity. *mBio*. 2014;5(4). <https://doi.org/10.1128/mbio.01371-14>.
23. Xiong C, He JZ, Singh BK, Zhu YG, Wang JT, Li PP, et al. Rare taxa maintain the stability of crop mycobiomes and ecosystem functions. *Environ Microbiol*. 2021;23(4):1907–24.
24. Leff JW, Jones SE, Prober SM, Barberán A, Borer ET, Firn JL, et al. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the Globe. *Proc Natl Acad Sci*. 2015;112(35):10967–72.
25. Gravuer K, Eskelinen A, Winbourne JB, Harrison SP. Vulnerability and resistance in the Spatial heterogeneity of soil microbial communities under resource additions. *Proc Natl Acad Sci*. 2020;117(13):7263–70.
26. Prosser JI. Replicate or lie. *Environ Microbiol*. 2010;12(7):1806–10.
27. Baker KL, Langenheder S, Nicol GW, Ricketts D, Killham K, Campbell CD, et al. Environmental and Spatial characterisation of bacterial community composition in soil to inform sampling strategies. *Soil Biol Biochem*. 2009;41(11):2292–8.
28. Child HT, Wierzbicki L, Joslin GR, Tennant RK. Comparative evaluation of soil DNA extraction kits for long read metagenomic sequencing. *Access Microbiol*. 2024;6(9):000868v3.
29. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. MetaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods*. 2020;17(11):1103–10.
30. Pan S, Zhao XM, Coelho LP. SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing. *Bioinformatics*. 2023;39(Supplement1):i21–9.
31. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods*. 2023;20(8):1203–12.
32. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J*. 2017;11(12):2864–8.
33. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. *Genome Biol*. 2019;20(1):257.
34. Huson DH, Beier S, Flade I, Górski A, El-Hadidi M, Mitra S et al. T PoisoT editor 2016 MEGAN community Edition - Interactive exploration and analysis of Large-Scale Microbiome sequencing data. *PLOS Comput Biol* 21;12(6):e1004957.
35. Dixon P. VEGAN, A package of R functions for community ecology. *J Veg Sci*. 2003;14(6):927–30.

Publisher's Note

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