

An ode to the 16S rRNA gene: its history, importance, caveats and future in microbiome research

F.J. Whelan^{1,*} and L.J. Hall^{2,3}

Abstract

The 16S rRNA gene has – and continues to – play an important role in microbiology. It's universality across prokaryotes and variation across species has allowed the sequencing of its hypervariable regions to be used to distinguish taxa within complex mixed bacterial communities. Although 16S rRNA gene sequencing has transformed our understanding of microbial communities, its use comes with important caveats and considerations, especially in light of the availability of whole-genome metagenomic sequencing. Within, we discuss the 16S rRNA gene and its important role in microbiome research, in the past, present, and future.

Impact Statement

The 16S rRNA gene has had a huge impact on the progress of microbiology and microbiome research. As new technologies such as whole-genome metagenomic sequencing become available, it is important for the community to consider the placement of 16S rRNA gene sequencing, its importance, and when it is most appropriate technology to use.

DATA SUMMARY

No new data were generated in this work. The data underlying Fig. 1 is publically available via PubMed and was plotted using open source software (<https://github.com/esperr/pubmed-by-year>).

ORIGINS AND USE: WHAT 16S rRNA GENE SEQUENCING GETS RIGHT

The past two decades have seen an explosion of research on microbial communities, in large part due to the widespread adoption of 16S rRNA gene sequencing. The use of the 16S rRNA gene for bacterial taxonomic identification and definition of phylogenetic relationships was pioneered in the 1970s by Carl Woese *et al.* [1, 2] and was advanced by Norman Pace and colleagues in the 1980s [3, 4] with the application to mixed microbial communities. Early studies involved labourious generation of clone libraries and subsequent Sanger sequencing [5]; however, uptake accelerated with next- (or second-) generation platforms, including 454 pyrosequencing [6] and Illumina paired-end reads [7] and now includes long-read sequencing [8].

The 16S rRNA gene was chosen because of its universality across prokaryotes, its high sequence conservation and the infrequency of its transfer horizontally between species [2, 3]. Furthermore, 16S rRNA's looped RNA structure of highly conserved stems and hypervariable loops meant that (almost) universal primers could be designed to the conserved gene regions to amplify the hypervariable loops. Generally, the sequences of the hypervariable regions of the 16S rRNA gene are unique to a given species or genus [9]; however, even full-length 16S rRNA gene sequences cannot always resolve taxa (e.g. differentiating *Shigella* species and *Escherichia coli*) [10, 11].

Received 23 January 2026; Accepted 03 August 2026; Published 25 August 2026

Author affiliations: ¹School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; ²Department of Microbes, Infection and Microbiomes, School of Infection, Inflammatory and Immunology, College of Medicine and Health, University of Birmingham, Birmingham, UK; ³Food, Microbiome and Health, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK.

***Correspondence:** F.J. Whelan, Fiona.whelan@manchester.ac.uk

Keywords: 16S rRNA gene sequencing; microbiome research.

Abbreviations: IBD, inflammatory bowel disease; ITS, internal transcribed spacer.

001823 © 2026 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

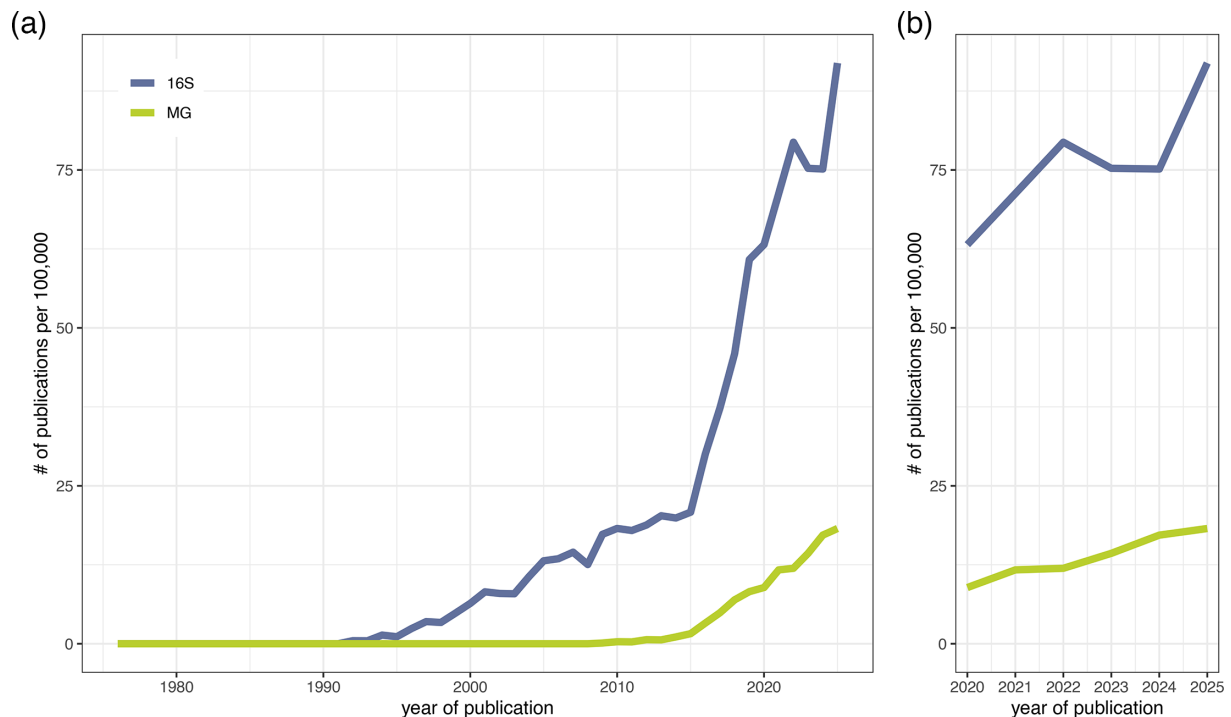


Fig. 1. Occurrences of the terms '16S rRNA gene sequencing' and 'shotgun metagenomics' in peer-reviewed manuscripts. Data were obtained from PubMed by year (<https://github.com/esperr/pubmed-by-year>) and plotted using R [52]. The data is based on the occurrence of each term in the full text of the paper since 1976 (a) and in the last 5 years (b). Data obtained 16 July 2025.

Platforms, such as Illumina's MiSeq originally offered 150bp paired-end reads, long enough to sequence some of the shorter variable regions (e.g. variable region 3, v3) or to obtain partial coverage of longer regions (e.g. v3–5) of the gene. The targeted variable region(s) differ across studies and large projects, including v1–3 and v3–5 in the Human Microbiome Project [12, 13], v4 in the Earth Microbiome Project [14] and v1–3 in the Microbial Database for Activated Sludge project [15]. Innovative techniques to 'barcode' the 16S rRNA gene amplicons from a given microbial community by the incorporation of unique 6-bp barcodes in the primer design [7] meant that many microbiome samples could be multiplexed onto a single Illumina sequencing run and, later, the reads separated by barcode.

The widespread adoption of high-throughput 16S rRNA gene sequencing has greatly enhanced our understanding of microbial communities. The impact has been felt in many diverse fields: 16S rRNA gene sequencing has predicted the ability of soil microbiomes to respond to climate change [16], indicated a 'core microbiome' of temperate trees [17] and a better understanding of 'culture-negative' bovine mastitis cases [18]. However, the place where 16S rRNA gene sequencing has perhaps been used most is in the study of the human microbiome. In this context, 16S rRNA gene amplicon sequencing has identified the influence of birth mode on the newborn microbiome [19], effects of antibiotics on the gut microbiome [20], the differences in microbial composition worldwide [21] and the identification of microbial communities in sites previously assumed sterile, such as the human lung [22, 23]. Community-level differences between health and disease have been commonly identified for gut diseases/disorders, such as inflammatory bowel disease (IBD) [24] and irritable bowel syndrome [25]. Additional studies have pinpointed particular changes in the relative abundance of key microbes and their correlation with human, animal and environmental health; for example, *Faecalibacterium prausnitzii* abundance is often reduced in individuals with various gut disorders [26]. None of these studies would have been possible without important innovations in bioinformatic and statistical analysis methods, including but certainly not limited to mothur [27], QIIME [28], DADA2 [29] and phyloseq [30].

CRITICAL CAVEATS (AND HOW TO ADDRESS THEM)

However, like all good things, 16S rRNA gene sequencing has important caveats that must be considered and accounted for. First, many microbial communities contain mixes of bacteria, archaea, fungi, viruses and other eukaryotic species, yet 16S rRNA gene sequencing is only able to capture the bacterial and archaeal diversity within these. Other amplicons – such as those targeting the 18S rRNA gene or internal transcribed spacer (ITS)1 and ITS2 – can help to elucidate some of these other members but others, such as bacteriophages and eukaryotic viruses, lack a universally conserved DNA target for amplicon sequencing. Additionally,

16S rRNA gene sequencing is inherently relative and cannot provide absolute counts. Community profiles can only indicate the proportional divisions of each taxon in the sample and are sensitive to microbial load. It can also only tell which bacteria are present but not their complement of genes – and thus, functionality. Furthermore, the ‘universal’ 16S rRNA gene primers developed for various hypervariable regions are not always quite as universal as hoped. For example, commonly used primers to amplify the v1–3 regions of the 16S rRNA gene have been shown to miss key health-associated genera, such as *Bifidobacterium* [31]. Further challenges in accurately interpreting the relative abundance of a species arise from differences in 16S rRNA gene sequence copy number (and, sometimes, the existence of distinct 16S rRNA genes within the same genome) [32, 33].

A critical caveat to consider when designing 16S rRNA gene sequencing studies is the potential for contamination [34]. Because 16S rRNA gene sequencing relies on amplification using primers universal to all prokaryotes, the method is prone to contamination. This is especially true of low-biomass samples where even low-level contamination can largely influence the resultant sequencing. Contamination can occur at multiple points from sample to sequence. First, contamination can be introduced during DNA extraction due to contaminants in commercial extraction kits and laboratory reagents [35]; in a seminal paper from 2014, the authors extracted DNA from a pure culture of *Salmonella bongori* at three different institutes, identifying at least 21 contaminating genera, the majority of which were consistent across laboratories, suggesting their presence in commercial reagents [35]. Most of these bacteria – for example *Hoeflea* spp. – are less commonly observed bacteria that wouldn’t be expected to be found in many niches (e.g. human gut) for which microbiome studies are conducted; however, others – for example, *Escherichia* spp. – are common members of many different environments and thus harder to distinguish from contaminant backgrounds [35]. Importantly, the relative abundance of contaminants observed increased as the amount of input *S. bongori* cells decreased, indicating the additional importance of mitigation against contamination when working with low-biomass samples [35]. Colloquially referred to as the ‘kitome’, the authors make recommendations on how to reduce the impact of contamination on low-biomass samples, which were later further outlined in opinion pieces [36] and consensus statements [37]. These recommendations include sequencing (a) sampling negatives (i.e. to control for contamination at the sampling source); (b) DNA extraction negatives (i.e. DNA extraction with no input sample) and positives (i.e. inclusion of a mock or synthetic community of known composition where contamination would be easy to identify); (c) PCR negatives (i.e. DNA extraction reagents) and positives (i.e. isolated DNA from a mock/synthetic community) [37]. Additionally, contamination can be mitigated by randomizing the extraction and sequencing of samples across treatment groups and/or other metadata variables and recording which samples were processed with which reagents and kits [35, 37]. Calls-to-action are being made to control for contamination in microbiome studies from environmental [37], clinical [38], animal [39] and insect [40] sources.

Despite best practices, contamination is still a real possibility. To help combat this are multiple *in silico* softwares which have been developed to aid in the removal of contaminant species. For example, Decontam is a statistical R package which identifies contamination based on increased abundance in low-concentration samples and presence in negative controls [41]. microDecon compares the proportion of potential contaminants in negative controls to sample datasets [42]. SourceTracker uses a Bayesian approach to determine potential contaminants and their source environments [43].

METAGENOMICS COMPLEMENTS – NOT REPLACES – AMPLICON SURVEYS

Despite these caveats, 16S rRNA gene sequencing has been used in thousands of peer-reviewed publications since the 1970s (Fig. 1a). However, in recent years, the use of 16S rRNA gene sequencing has been plateauing alongside an increase in the use of whole-genome metagenomic sequencing (Fig. 1a and 1b). Metagenomics – the sequencing of the totality of the DNA present in a given sample [44] – is the natural next step from amplicon sequencing and addresses some of the caveats of 16S rRNA gene amplicons. Originally defined in 1998 [44] and used in 2002 [45], metagenomics is not limited to identifying only bacterial DNA but sequences the DNA of all organisms present in a sample. For communities, such as the human gut microbiome with its known diversity of fungi, viruses and free-living phage [46], metagenomics has allowed further characterization of these organisms as part of these mixed microbial communities. Metagenomics also allows better functional inference of the community by capturing the genomic content of at least the highly abundant organisms (dependent on the sequencing depth used and amount of non-microbial DNA present). Metagenomics does, however, still suffer from similar issues of contamination in low-biomass samples and community composition still being reported as ratios and not absolute counts.

Despite the wide-scale recent adoption of metagenomic sequencing, we argue that 16S rRNA gene sequencing is still a useful methodology and should be considered the technology of choice depending on the community being studied and the scientific question being asked. First, 16S rRNA gene amplicons are a relatively inexpensive way of identifying who is present in a community and how that presence might change over time, individual, disease state, etc. Its low cost means that sample sets containing thousands of samples can be assessed where similar numbers of shotgun metagenomics might be beyond the research budget or computational/bioinformatic capacity. It is unhelpful, for example, as a reviewer to suggest that metagenomics must be used if the research question can be addressed with a cheaper, well-established technology. Second, the technology continues to improve, including improvements to long-read 16S rRNA gene sequencing which allows for full-length sequencing of the gene, making this technology powerful for genus- (and sometimes species-) level resolution [8]. Third, metagenomics may have limited use

in communities with large amounts of human/host DNA. One of the advantages of metagenomics – its representation of the DNA from all organisms in a community – can also be its disadvantage in communities where the vast majority of that DNA is non-microbial. For example, metagenomic studies of the cystic fibrosis human lung microbiomes can result in up to 99% of human DNA in their raw sequencing [47–50]. Not only does this introduce ethical concerns [51] but it also means that 99% of the output of an already relatively more expensive technique must then be discarded.

CONCLUSIONS

16S rRNA gene sequencing was first used by Woese to identify a new kingdom of life before going on to aid in the discovery of the role of the microbiome in gut diseases, such as IBD and Crohn's. It has uncovered the unimaginable diversity of assorted environments and continues to be a useful technology for specific, focused research questions. We don't think this influential gene – when used appropriately – has quite finished helping us reveal the magical world of microbial communities just yet.

Funding information

F.J.W. acknowledges funding from a UKRI Future Leaders Fellowship (MR/Y016343/1), AMS Springboard (SBF009\1062) and the Lister Institute Research Prize. L.J.H. is supported by a Wellcome Trust Investigator Award no. 220876/Z/20/Z and a Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Programme Food Microbiome and Health BB/X011054/1 and its constituent project BBS/E/F/000PR1361.

Acknowledgements

The authors thank Michael G Surette, Jennifer Lau, Louis P Schenck and Daphnée Lamarche for conversations critical to this work.

Author contributions

F.J.W.: Conceptualization, data curation, funding acquisition, visualization, writing – original draft, writing – review and editing. L.J.H.: Conceptualization, writing – original draft, writing – review and editing.

Conflicts of interest

The author(s) declare that there are no conflicts of interest.

1. Woese CR, Fox GE, Zablen L, Uchida T, Bonen L, *et al.* Conservation of primary structure in 16S ribosomal RNA. *Nature* 1975;254:83–86.
2. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 1977;74:5088–5090.
3. Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA. Microbial ecology and evolution: a ribosomal RNA approach. *Annu Rev Microbiol* 1986;40:337–365.
4. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, *et al.* Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* 1985;82:6955–6959.
5. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 1990;345:60–63.
6. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, *et al.* Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* 2006;103:12115–12120.
7. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl Environ Microbiol* 2011;77:3846–3852.
8. Aja-Macaya P, Conde-Pérez K, Trigo-Tasende N, Buetas E, Nasser-Ali M, *et al.* Nanopore full length 16S rRNA gene sequencing increases species resolution in bacterial biomarker discovery. *Sci Rep* 2025;15:26486.
9. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 1991;173:697–703.
10. Khot PD, Fisher MA. Novel approach for differentiating *Shigella* species and *Escherichia coli* by matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *J Clin Microbiol* 2013;51:3711–3716.
11. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY, Veeraraghavan B. Accurate differentiation of *Escherichia coli* and *Shigella* serogroups: challenges and strategies. *New Microbes New Infects* 2018;21:58–62.
12. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* 2012;486:215–221.
13. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human. *Nature* 2012;207–214.
14. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* 2011;108 Suppl 1:4516–4522.
15. McIlroy SJ, Saunders AM, Albertsen M, Nierychlo M, McIlroy B, *et al.* MiDAS: the field guide to the microbes of activated sludge. *Database (Oxford)* 2015;2015:bav062.
16. Knight CG, Nicolitch O, Griffiths RI, Goodall T, Jones B, *et al.* Soil microbiomes show consistent and predictable responses to extreme events. *Nature* 2024;636:690–696.
17. Laforest-Lapointe I, Messier C, Kembel SW. Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. *Microbiome* 2016;4:27.
18. Kuehn JS, Gorden PJ, Munro D, Rong R, Dong Q, *et al.* Bacterial community profiling of milk samples as a means to understand culture-negative bovine clinical mastitis. *PLoS One* 2013;8:e61959.
19. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA* 2010;107:11971–11975.
20. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* 2008;6:e280.
21. Abdill RJ, Graham SP, Rubinetti V, Ahmadian M, Hicks P, *et al.* Integration of 168,000 samples reveals global patterns of the human gut microbiome. *Cell* 2025;188:1100–1118.
22. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, *et al.* Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011;184:957–963.
23. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, *et al.* Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med* 2013;187:1067–1075.
24. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, *et al.* Molecular-phylogenetic characterization of microbial community

- imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* 2007;104:13780–13785.
25. Rajilić-Stojanović M, Biagi E, Heilig HGHJ, Kajander K, Kekkonen RA, *et al*. Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology* 2011;141:1792–1801.
 26. Lopez-Siles M, Duncan SH, Garcia-Gil LJ, Martinez-Medina M. *Faecalibacterium prausnitzii*: from microbiology to diagnostics and prognostics. *ISME J* 2017;11:841–852.
 27. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, *et al*. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537–7541.
 28. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, *et al*. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–336.
 29. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, *et al*. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13:581–583.
 30. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;8:e61217.
 31. Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, *et al*. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome* 2015;3:26.
 32. Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, *et al*. Diversity of 16S rRNA genes within individual prokaryotic genomes. *Appl Environ Microbiol* 2010;76:3886–3897.
 33. Větrovský T, Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 2013;8:e57923.
 34. Tanner MA, Goebel BM, Dojka MA, Pace NR. Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl Environ Microbiol* 1998;64:3110–3113.
 35. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, *et al*. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014;12:87.
 36. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, *et al*. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol* 2019;27:105–117.
 37. Fierer N, Leung PM, Lappan R, Eisenhofer R, Ricci F, *et al*. Guidelines for preventing and reporting contamination in low-biomass microbiome studies. *Nat Microbiol* 2025;10:1570–1580.
 38. Dyrhovden R, Rippin M, Øvrebo KK, Nygaard RM, Ulvestad E, *et al*. Managing contamination and diverse bacterial loads in 16S rRNA deep sequencing of clinical samples: implications of the law of small numbers. *mBio* 2021;12:e0059821.
 39. Weinroth MD, Belk AD, Dean C, Noyes N, Dittoe DK, *et al*. Considerations and best practices in animal science 16S ribosomal RNA gene sequencing microbiome studies. *J Anim Sci* 2022;100:skab346.
 40. Williamson EM, Hammer TJ, Hogendoorn K, Eisenhofer R. Blanking on blanks: few insect microbiota studies control for contaminants. *mBio* 2025;16:e0265824.
 41. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 2018;6:226.
 42. McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, *et al*. microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. *Environmental DNA* 2019;1:14–25.
 43. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, *et al*. Bayesian community-wide culture-independent microbial source tracking. *Nat Methods* 2011;8:761–763.
 44. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 1998;5:R245–9.
 45. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, *et al*. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* 2002;99:14250–14255.
 46. Shanahan F, Ghosh TS, O'Toole PW. The healthy microbiome—what is the definition of a healthy gut microbiome? *Gastroenterology* 2021;160:483–494.
 47. Whelan FJ, Waddell B, Syed SA, Shekarriz S, Rabin HR, *et al*. Culture-enriched metagenomic sequencing enables in-depth profiling of the cystic fibrosis lung microbiota. *Nat Microbiol* 2020;5:379–390.
 48. Lim YW, Schmieder R, Haynes M, Willner D, Furlan M, *et al*. Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. *J Cyst Fibros* 2013;12:154–164.
 49. Huang YJ, LiPuma JJ. The microbiome in cystic fibrosis. *Clin Chest Med* 2016;37:59–67.
 50. Zhao J, Schloss PD, Kalikin LM, Carmody LA, Foster BK, *et al*. Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc Natl Acad Sci USA* 2012;109:5809–5814.
 51. Sarhan MS, Antonello G, Weissensteiner H, Mengoni C, Mascalzoni D, *et al*. Human mitochondrial DNA in public metagenomes: opportunity or privacy threat? *Cell* 2025;188:2561–2566.
 52. The R Foundation. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2025.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org