Review Article

Next-Generation Metagenomics: Methodological Challenges and Opportunities

AU2 ▶

AU3 ▶

Ilaria Laudadio,* Valerio Fulci,* Laura Stronati, and Claudia Carissimi

Abstract

Metagenomics is not only one of the newest omics system science technologies but also one that has arguably the broadest set of applications and impacts globally. Metagenomics has found vast utility not only in environmental sciences, ecology, and public health but also in clinical medicine and looking into the future, in planetary health. In line with the One Health concept, metagenomics solicits collaboration between molecular biologists, geneticists, microbiologists, clinicians, computational biologists, plant biologists, veterinarians, and other health care professionals. Almost every ecological niche of our planet hosts an extremely diverse community of organisms that are still poorly characterized. Detailed characterization of the features of such communities is instrumental to our comprehension of ecological, biological, and clinical complexity. This expert review article evaluates how metagenomics is improving our knowledge of microbiota composition from environmental to human samples. Furthermore, we offer an analysis of the common technical and methodological challenges and potential pitfalls arising from metagenomics approaches, such as metagenomics study design, data processing, and interpretation. All in all, at this critical juncture of further growth of the metagenomics field, it is time to critically reflect on the lessons learned and the future prospects of next-generation metagenomics science, technology, and conceivable applications, particularly from the standpoint of a metagenomics methodology perspective.

Keywords: metagenomics, one health, next-generation sequencing, ecology, environmental science, biomarkers, microbial genomics

Introduction

DURING EVOLUTION, MICROORGANISMS HAVE ADAPTED TO an incredibly diverse set of environments. Almost every ecological niche of our planet hosts an extremely diverse community of unicellular organisms that are still poorly characterized. Detailed depiction of the features of such communities is instrumental to our comprehension of ecological, biological, and clinical issues. The study of the collective genome of microorganisms from an environmental sample is named metagenomics (Xia et al., 2018).

Metagenomic analyses of soil and marine ecosystems have proven that a hitherto unappreciated genetic diversity exists. Both soil (Daniel, 2005; Thompson et al., 2017) and oceanic (Bork et al., 2015; Rusch et al., 2007; Yooseph et al., 2007) metagenomes were investigated through various global sample collection efforts, unveiling unexpected diversities in prokaryotic, unicellular eukaryotic (Carradec et al., 2018) organisms and viruses (Schulz et al., 2018). These efforts yielded a large amount of genetic data that can be explored for ecological, biotechnology, and phylogenetic applications, among others. Environmental metagenomics can be exploited to assess the effects of contamination by different pollutants on microbial communities (Ghosh and Das, 2018; Hemme et al., 2015; Jung et al., 2016), thus suggesting possible bioremediation strategies. Metagenomics has also been harnessed so as to investigate the microbiomes associated with agriculture and food (Liu et al., 2018; Orellana et al., 2018).

Mammals harbor complex communities of microorganisms that live on the surface and inside the host (skin, mucosal surfaces, gut, urogenital system, etc.). The mammalian microbiota includes all the domains of life (Archaea, Bacteria, and Eukaryota) and is fundamental to the health of the host (D'Argenio and Salvatore, 2015). Studies in mammals have implicated the microbial communities present in the gastrointestinal tract (gut microbiota) in a range of physiologic processes with a significant impact on food digestion, metabolic processes, immunity, acquisition, and maintenance of overall wellness of the host. Accordingly, mounting evidence suggests a correlation between changes in composition of the

AU4 ▶

AU5 ▶

Department of Molecular Medicine, "Sapienza" University of Rome, Rome, Italy.
*These authors contributed equally to the article as first authors.

gut microbiota and the pathophysiology of several disorders such as depression, autism, allergies, and chronic inflammatory diseases (Fazlollahi et al., 2018; Haberman et al., 2014; Hsiao et al., 2013; Imhann et al., 2018; Savage et al., 2018; Sekirov et al., 2010; Trompette et al., 2014).

Metagenomics is likely to have transformative and possibly game-changing impacts in clinical practice as well. On one hand, a microbial dysbiosis could be causally linked to a disease, suggesting targets of new therapeutic interventions aiming at dysbiosis correction and restoration of the so-called eubiosis. On the other hand, metagenomic profile could also represent a noninvasive, cost-effective, and rapid biomarker useful for diagnosis and/or prognosis. In infectious diseases, prediction of antimicrobial resistance adds a further layer of relevant information that can be retrieved from clinical specimens (Wilson et al., 2014). Furthermore, these approaches also impact public health with applications such as monitoring antimicrobial resistance in food supply by bacterial whole-genome sequencing (Chiu and Miller, 2019; Oniciuc et al., 2018).

Despite its relatively brief history, the study of microbial life through next-generation sequencing (NGS) technologies and computational biology is defining a new era in microbiology. Up to now, two main NGS-based strategies have been implemented for whole microbial genome analysis: 16S ribosomal DNA (rDNA) sequencing and shotgun metagenomics. The former relies on PCR amplification of hypervariable regions of bacterial 16S rDNA through the use of degenerated primers followed by amplicon deep sequencing. Intrinsic limits of this approach are the exclusion of eukaryotic organisms and viruses from the analysis and the possible primer biases toward specific taxa (Tremblay et al., 2015). Since 16S amplicon sequencing focuses on the analysis of a tiny region of prokaryotic DNA, it is not formally defined as metagenomics (Xia et al., 2018). Differently, shotgun metagenomics (or simply metagenomics) relies instead on sequencing of randomly sampled DNA fragments isolated from a microorganism community. Metagenomics allows the characterization of complex communities of microorganisms of specific environments, including human body sites, at unprecedented resolution, without the need of any *a priori* knowledge and without prior culturing (Escobar-Zepeda et al., 2015; Schloss and Handelsman, 2005; Tringe et al., 2005).

Several reports compared the performance of the two techniques. Comparison of 16S rDNA sequencing and metagenomics has been performed using a single sample on Solid (Mitra et al., 2013) and Illumina platforms (Ranjan et al., 2016). The use of synthetic samples has also been exploited to compare the two approaches (Jovel et al., 2016). Such comparisons showed that the 16S rDNA amplicon sequencing approach yields quantitatively and qualitatively different results compared to metagenomics. In a recent article, we compared the performance of metagenomics and 16S amplicon sequencing using Illumina platform by using DNA isolated from human fecal samples, showing that metagenomics outperforms 16S rDNA amplicon sequencing (Laudadio et al., 2018).

This expert review article evaluates how metagenomics is improving our knowledge of microbiota composition from environmental to human samples. Furthermore, we offer an analysis of the common technical and methodological challenges and potential pitfalls arising from metagenomics approaches, such as metagenomics study design, data processing, and interpretation.

Current Approaches and Future Challenges in Metagenomics

From experimental design to genome assembly

A major bottleneck of metagenomics NGS is the ability to translate the data into relevant information, to obtain clinically actionable results. A professional biostatistician needs to be consulted at the time of study design (before samples collection) to programmatically assess statistical power to be achieved and which metadata should be collected and included in the analysis (Fig. 1). Overlooking the relevance of ◀F1 this step will invariably lead to difficulties in the subsequent analysis of data and negatively affect the relevance of any finding. Furthermore, confounding factors such as diet, environment, and social behavior, which may affect human microbiota, have to be considered. To avoid biases due to cultural or social issues, it is recommended that broad cohorts of patients (and controls) are analyzed, possibly through large
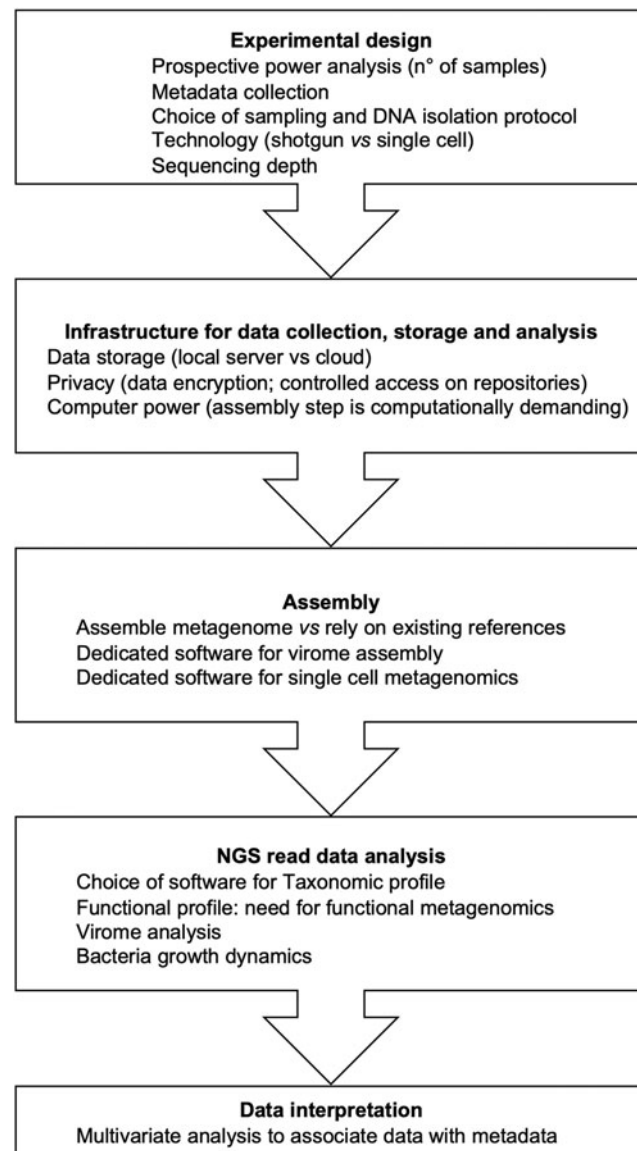

Experimental design
Prospective power analysis (n° of samples)
Metadata collection
Choice of sampling and DNA isolation protocol
Technology (shotgun *vs* single cell)
Sequencing depth

Infrastructure for data collection, storage and analysis
Data storage (local server vs cloud)
Privacy (data encryption; controlled access on repositories)
Computer power (assembly step is computationally demanding)

Assembly
Assemble metagenome *vs* rely on existing references
Dedicated software for virome assembly
Dedicated software for single cell metagenomics

NGS read data analysis
Choice of software for Taxonomic profile
Functional profile: need for functional metagenomics
Virome analysis
Bacteria growth dynamics

Data interpretation
Multivariate analysis to associate data with metadata

FIG. 1. Typical metagenomic study workflow: key choices and critical parameters.

AU11 ▶ TABLE 1. HOW METAGENOMICS CHANGED OUR APPROACH TO MICROBIOLOGY AND FUTURE DIRECTIONS

State of the art metagenomics	Next generation metagenomics
Discovery of novel species and strains	Characterization of variability within each species and strain through single-cell metagenomics
Taxonomic and functional community profiles associated with health/disease or environmental changes	Mechanistic insights into the role of specific microorganisms in health and disease or in specific ecosystems
Ongoing annotation of genomes and genes	Exhaustive metagenomic databases; extensive characterization of microbial gene function
Computationally intensive assembly of metagenomes from shotgun DNA fragments arising from different cells	Assembly of genomes from single-cell shotgun metagenomics

consortia, allowing to integrate in a single study cohort of patients from different continents.

One of the most relevant constraints of metagenomics lies in the limited annotation of bacterial genomes. Human gut microbiome metagenomics took advantage of seminal studies, which allowed a very deep coverage on hundreds to thousands of samples and resulted in comprehensive coverage and extensive annotation of human gut bacterial genomes (Li et al., 2014; Pasolli et al., 2019; The Integrative Human Microbiome Project, 2014). On the other hand, metagenomes from poorly exploited environmental niche are still inadequately characterized (Uritskiy and DiRuggiero, 2019); hence, databases offer only limited coverage for alignment of reads. In these instances, the first challenging step is the assembly of a reference metagenome to use in the analysis.

Assembly is a key step of the analysis pipeline as all subsequent output will depend on its outcome. Some features of metagenomic data (i.e., uneven coverage across species, highly similar sequences in unrelated species due to horizontal gene transfer, and repeated sequences) require dedicated software. Binning co-abundant sequences before assembly has been proven to overcome limitations due to uneven coverage (Plaza Oñate et al., 2018). Several different strategies can be exploited (Ghurye et al., 2016), and a wide array of software to assemble metagenomes have been developed and tested on both real samples (Olson et al., 2017; Wang et al., 2019) and synthetic bacterial communities (Greenwald et al., 2017). It has been shown that virome analysis (a branch of metagenomics focusing on bacteriophage communities) is particularly affected by the choice of the assembler software used (Sutton et al., 2019).

Metagenomics analysis: one raw reads dataset, several layers of information

Through the use of several available tools, bioinformatics analysis of metagenomics datasets can provide information at different levels. Even though, metagenomics datasets are most commonly investigated to achieve information on the taxonomic and functional profiles of a microbiome, tools to investigate the virome (Nooij et al., 2018; Ogilvie and Jones, 2015; Rampelli et al., 2016), the replication rates of bacteria (Korem et al., 2015), and the profile of antibiotic resistance (Rowe and Winn, 2018; Yang et al., 2016) have also been set up. In particular, virome investigation through metagenomics has revolutionized the field of virology, disclosing a wealth of putative novel viruses (Schulz et al., 2018; Simmonds et al., 2017). Moreover, bacteriophages have recently emerged as key remodelers of microbial host communities, influencing the bacterial diversity, facilitating nutrient turnover, and conferring antibiotic resistance genes through horizontal transfer of genetic material (Modi et al., 2013; Ogilvie and Jones, 2015; Reyes et al., 2013).

Taxonomic profiles are obtained by mapping reads on a database of reference bacterial genomes. Most software rely on a small selection of genes (markers) to obtain a taxonomic profile (Nayfach et al., 2016; Segata et al., 2012; Truong et al., 2017). The choice of the selection of marker genes allows users to obtain taxonomic profiles with relatively low computing power.

Of note, contamination-free collection, homogenization, storage, and a subsequent efficient DNA extraction method have been shown to significantly affect the taxonomic profile yielded (Brooks et al., 2015; Wesolowska-Andersen et al., 2014). Several studies documented that contaminant DNA and cross-contamination can critically influence NGS-based microbiome analyses (Salter et al., 2014; Sinha et al., 2015). Contaminant DNA appears to originate from reagents, laboratory environments, human commensals on laboratory personnel, and sample processing. Indeed, for the Human Microbiome Project, a rigorous study protocol and standardized instructions for body site sampling and specimen processing were set up (Aagaard et al., 2013). Moreover, Panek et al. (2018) described the influence of storage condition and sample extraction in detection and composition of the fecal bacterial community.

Functional metagenomics: from microbial community biodiversity to functional processes

Functional profiling aims at achieving a comprehensive view of the functions of proteins encoded by the metagenome. Such analysis is extremely relevant for both ecologic and clinic purposes, as it highlights the interplay between a specific microbial community and the environment. Metabolomic analyses are the natural complement of a functional profiling, allowing the researcher to correlate alteration in the abundance of specific metabolites and that of taxa and genes related to those metabolites (Frankel et al., 2017; Liu et al., 2017).

16S amplicon sequencing allowed functional profiling by assuming that it is the sum of the functional profiles of all the identified species (Aßhauer et al., 2015; Langille et al., 2013). Otherwise, metagenomics yields functional profiles through direct identification of the genes encoded by a metagenome rather than inferring them based on taxonomy. Furthermore, most pipelines can assign a function not only to reads mapping to annotated genes but also to a fraction of the reads that cannot be precisely assigned to an annotated genome. This is exerted through *in silico* translation of the

DNA sequences, followed by homology search on protein sequence databases. This process, called ''translated search,'' allows the software to ascribe a putative function to the protein encoded by a DNA fragment (Buchfink et al., 2015; Franzosa et al., 2018; Huson et al., 2016). While extensive annotation of the gut microbiome allows binning of most reads in gut microbiota, in environmental metagenomics, most reads cannot be assigned to any function, due to lack of annotation and poor knowledge of the function of most polypeptides encoded by microorganisms (Quince et al., 2017), reflecting the current bias of the annotation toward a small number of cultivable microorganisms. Functional metagenomics, aiming at isolation, cloning, and expressing genes into suitable model organisms to identify their function will be required to achieve a comprehensive annotation of microbial genes (Santana-Pereira and Liles, 2017), thus improving the resolution of functional profiles obtained through metagenomics.

An extremely relevant subfield of functional metagenomics is related to antibiotic resistance, aiming at the identification of genes that confer antibiotic resistance to microorganisms. By leveraging information collected in this field, it is possible to characterize the ''resistome'' in both environmental (D'Costa et al., 2006) and human samples (van Schaik Willem, 2015) with relevant applications in built environments such as hospitals (Mahnert et al., 2019).

AU6 ▶

In the past decade, transformation of biomedical big data into valuable knowledge has been a fundamental challenge in bioinformatics, and a significant role in the success of bioinformatics exploration from biological data came from machine learning and deep learning applications. The intrinsic characteristics of several steps of metagenomic analyses are suitable of being addressed through machine learning and neural networks. In fact, such methods have been exploited for metagenome assembly (Afiahayati et al., 2015; Ji et al., 2017; Wang et al., 2015), metagenomic reads binning to reference genomes (Vervier et al., 2018), gene prediction (Al-Ajlan and El Allali, 2018), gene function prediction (Li et al., 2018), and analyses of large datasets (LaPierre et al., 2019; Pasolli et al., 2016).

Big data storage and analysis considerations

The large amount of data typically generated by metagenomic opens two major challenges: data storage and computational power. While some tasks such as taxonomic analysis with marker genes can be achieved with limited resources (assuming that an appropriate reference metagenome is already available), more complex tasks (metagenome assembly, functional profiling, and gene discovery) are much more demanding in terms of computational resources. Cloud computing can certainly provide an affordable solution; indeed, several server implementations of metagenomics pipelines are available (Lee et al., 2018; Mitchell et al., 2018; Raknes and Bongo, 2018), while other packages are released as images installable on cloud computing servers (McIver et al., 2018). However, privacy concerns may arise when human microbiome data are being analyzed and stored. In fact, several seminal studies highlighted that precise identification of individuals by means of metagenomics is feasible (Franzosa et al., 2015; Leake et al., 2016; Schmedes et al., 2017). Therefore, this issue started to be addressed by

implementation of metagenomics analysis using secure computation (Wagner et al., 2016).

Conclusions and Future Outlook

Shotgun metagenomics has rapidly expanded our understanding of environmental and clinical microbial communities, allowing to address previously unattainable biological questions as well as accelerating genome-based discovery of novel microbial genes (Santana-Pereira and Liles, 2017). One of the most intriguing fields propelled by metagenomics data will be functional metagenomics. The development of medium- to high-throughput tools to characterize the thousands of novel genes encoded by bacterial and viral genomes will likely result in a collection of exploitable enzymes, whose applications will range from biotechnology to material science and pharmacology. Investigation of the secondary metabolites produced by bacteria will also represent an invaluable resource.

In the last decade, according to PubMed data (Mesh terms search), 4227 scientific reports have been published on ''metagenomics,'' 1163 of which attain the ''microbiome, human.'' Indeed, characterization of the human microbiome unveiled that microbial communities associated with human body sites are dynamic and subject to impressive modification in the course of host life and in response to many factors, including the interaction between human, animal, and environment. Changes in the microbiome have been associated with disease states, and in some instances, causal links were highlighted between microbiota and specific pathological conditions. Accordingly, the ''One Health'' concept underlines the inseparable ecological relationships between human, animal, and environmental, recognizing that the health of people is connected to the health of animals and the environment. Indeed, the ''One Health'' Commission declared that nearly 75% of emerging human infectious diseases in the past three decades originated in animals (Murtaugh et al., 2017). The development of high-throughput tools to characterize entire microbial communities now offers new insights toward ''One Health'' concept. Actually, metagenomics approach will integrate the knowledge of the complex interactions from these three domains, open the potential for novel diagnostic tools, and pave the way to collaborative and integrated multidisciplinary approaches to treatment and intervention.

Most recent technological developments allowed single-cell metagenomics (Stepanauskas and Sieracki, 2007; Stepanauskas et al., 2017). This innovative approach overcomes some relevant limits intrinsic to the shotgun sequencing, as assembly of a genome from fragments arising from a single cell is computationally much less demanding. In fact, while metagenome assembly from shotgun sequencing requires multiple comparisons between many fragments, which in most cases arise from different genomes, single-cell metagenomics assembly takes advantage of smaller datasets of DNA reads, all arising from a single cell, allowing also a much easier parallelization of the assembly process. Furthermore, comparison of the two methods not only provided an implicit independent confirmation of the current metagenomics assembly methods but also highlighted a thus far overlooked variability between single cells belonging to the same taxon (Alneberg et al., 2018). Nevertheless, further improvements and widespread application of single-cell metagenomics are required to fully unleash the potential of this approach.

Many high-throughput techniques have been developed through step-wise improvements and have required, at some point, a standardization effort to allow proper comparison of data produced by different groups (Brazma et al., 2001; Taylor et al., 2007, 2008). The metagenomics community is currently setting up reliable and consistent standard procedures for sample collection, storage, and processing, data analysis, and metagenome reporting (Bowers et al., 2017; Roux et al., 2018). Further effort to standardize metagenomics database will be necessary to support consistent and productive development of the field.

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

Aagaard K, Petrosino J, Keitel W, et al. (2013). The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. FASEB J Off Publ Fed Am Soc Exp Biol 27, 1012–1022.

Afiahayati, Sato K, and Sakakibara Y. (2015). MetaVelvet-SL: An extension of the Velvet assembler to a de novo metagenomic assembler utilizing supervised learning. DNA Res 22, 69–77.

Al-Ajlan A, and El Allali A. (2018). CNN-MGP: Convolutional neural networks for metagenomics gene prediction. Interdiscip Sci Comput Life Sci [Epub ahead of print]; DOI: 10.1007/s12539-018-0313-4.

Alneberg J, Karlsson CMG, Divne A-M, et al. (2018). Genomes from uncultivated prokaryotes: A comparison of metagenome-assembled and single-amplified genomes. Microbiome 6, 173.

Aßhauer KP, Wemheuer B, Daniel R, et al. (2015). Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31, 2882–2884.

Bowers RM, Kyrpides NC, Stepanauskas R, et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35, 725–731.

Brazma A, Hingamp P, Quackenbush J, et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29, 365–371.

Brooks JP, Edwards DJ, Harwich MD, et al. (2015). The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies. BMC Microbiol 15, 66.

Buchfink B, Xie C, and Huson DH. (2015). Fast and sensitive protein alignment using DIAMOND. Nat Methods 12, 59–60.

Carradec Q, Pelletier E, Silva CD, et al. (2018). A global ocean atlas of eukaryotic genes. Nat Commun 9, 373.

AU7 ▶ Chiu CY, and Miller SA. (2019). Clinical metagenomics. Nat Rev Genet.

Daniel R. (2005). The metagenomics of soil. Nat Rev Microbiol 3, 470–478.

D'Argenio V, and Salvatore F. (2015). The role of the gut microbiome in the healthy adult status. Clin Chim Acta Int J Clin Chem 451, 97–102.

Escobar-Zepeda A, Vera-Ponce de León A, and Sanchez-Flores A. (2015). The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. Front Genet 6, 348.

Fazlollahi M, Chun Y, Grishin A, et al. (2018). Early-life gut microbiome and egg allergy. Allergy 73, 1515–1524.

Frankel AE, Coughlin LA, Kim J, et al. (2017). Metagenomic shotgun sequencing and unbiased metabolomic profiling identify specific human gut microbiota and metabolites associated with immune checkpoint therapy efficacy in melanoma patients. Neoplasia NYN 19, 848–855.

Franzosa EA, Huang K, Meadow JF, et al. (2015). Identifying personal microbiomes using metagenomic codes. Proc Natl Acad Sci U S A 112, E2930–E2938.

Franzosa EA, McIver LJ, Rahnavard G, et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods 15, 962–968.

Ghosh S, and Das AP. (2018). Metagenomic insights into the microbial diversity in manganese-contaminated mine tailings and their role in biogeochemical cycling of manganese. Sci Rep 8, 8257.

Ghurye JS, Cepeda-Espinoza V, and Pop M. (2016). Metagenomic assembly: Overview, challenges and applications. Yale J Biol Med 89, 353–362.

Greenwald WW, Klitgord N, Seguritan V, et al. (2017). Utilization of defined microbial communities enables effective evaluation of meta-genomic assemblies. BMC Genomics 18, 296.

Haberman Y, Tickle TL, Dexheimer PJ, et al. (2014). Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. J Clin Invest 124, 3617–3633.

Hemme CL, Tu Q, Shi Z, et al. (2015). Comparative metagenomics reveals impact of contaminants on groundwater microbiomes. Front Microbiol 6, 1205.

Hsiao EY, McBride SW, Hsien S, et al. (2013). Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell 155, 1451–1463.

Huson DH, Beier S, Flade I, et al. (2016). MEGAN community edition—Interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol 12, e1004957.

Imhann F, Vich Vila A, Bonder MJ, et al. (2018). Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. Gut 67, 108–119.

Integrative HMP (iHMP) Research Network Consortium. (2014). The Integrative Human Microbiome Project: Dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe 16, 276–289.

Ji P, Zhang Y, Wang J, et al. (2017). MetaSort untangles metagenome assembly by reducing microbial community complexity. Nat Commun 8, 14306.

Jovel J, Patterson J, Wang W, et al. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. Front Microbiol 7, 459.

Jung J, Philippot L, and Park W. (2016). Metagenomic and functional analyses of the consequences of reduction of bacterial diversity on soil functions and bioremediation in diesel-contaminated microcosms. Sci Rep 6, 23012.

Korem T, Zeevi D, Suez J, et al. (2015). Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. Science 349, 1101–1106.

Langille MGI, Zaneveld J, Caporaso JG, et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31, 814–821.

LaPierre N, Ju CJ-T, Zhou G, et al. (2019). MetaPheno: A ◀ AU8 critical evaluation of deep learning and machine learning in metagenome-based disease prediction. Methods.

Laudadio I, Fulci V, Palone F, et al. (2018). Quantitative assessment of shotgun metagenomics and 16S rDNA amplicon sequencing in the study of human gut microbiome. Omics J Integr Biol 22, 248–254.

Leake SL, Pagni M, Falquet L, et al. (2016). The salivary microbiome for differentiating individuals: Proof of principle. Microbes Infect 18, 399–405.

Lee B, Min H, and Yoon S. (2018). MUGAN: Multi-GPU accelerated AmpliconNoise server for rapid microbial diversity assessment. Bioinformatics [Epub ahead of print]; DOI: 10.1093/bioinformatics/bty096.

Li J, Jia H, Cai X, et al. (2014). An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 32, 834–841.

Li Y, Wang S, Umarov R, et al. (2018). DEEPre: Sequence-based enzyme EC number prediction by deep learning. Bioinformatics 34, 760–769.

Liu J, Cade-Menun BJ, Yang J, et al. (2018). Long-term land use affects phosphorus speciation and the composition of phosphorus cycling genes in agricultural soils. Front Microbiol 9, 1643.

Liu R, Hong J, Xu X, et al. (2017). Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. Nat Med 23, 859–868.

Mahnert A, Moissl-Eichinger C, Zojer M, et al. (2019). Man-made microbial resistances in built environments. Nat Commun 10, 968.

McIver LJ, Abu-Ali G, Franzosa EA, et al. (2018). bioBakery: A meta'omic analysis environment. Bioinformatics 34, 1235–1237.

Mitchell AL, Scheremetjew M, Denise H, et al. (2018). EBI metagenomics in 2017: Enriching the analysis of microbial communities, from sequence reads to assemblies. Nucleic Acids Res 46, D726–D735.

Mitra S, Förster-Fromme K, Damms-Machado A, et al. (2013). Analysis of the intestinal microbiota using SOLiD 16S rRNA gene sequencing and SOLiD shotgun sequencing. BMC Genomics 14, S16.

Modi SR, Lee HH, Spina CS, et al. (2013). Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. Nature 499, 219–222.

Murtaugh MP, Steer CJ, Sreevatsan S, et al. (2017). The science behind One Health: At the interface of humans, animals, and the environment. Ann NY Acad Sci 1395, 12–32.

Nayfach S, Rodriguez-Mueller B, Garud N, et al. (2016). An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res 26, 1612–1625.

Nooij S, Schmitz D, Vennema H, et al. (2018). Overview of virus metagenomic classification methods and their biological applications. Front Microbiol 9, 749.

Ogilvie LA, and Jones BV. (2015). The human gut virome: A multifaceted majority. Front Microbiol 6, 918.

Olson ND, Treangen TJ, Hill CM, et al. (2017). Metagenomic assembly through the lens of validation: Recent advances in assessing and improving the quality of genomes assembled from metagenomes. Brief Bioinform [Epub ahead of print]; DOI: 10.1093/bib/bbx098.

Oniciuc EA, Likotrafiti E, Alvarez-Molina A, et al. (2018). The present and future of whole genome sequencing (WGS) and whole metagenome sequencing (WMS) for surveillance of antimicrobial resistant microorganisms and antimicrobial resistance genes across the food chain. Genes 9, pii: E268.

AU9 ► Orellana LH, Chee-Sanford JC, Sanford RA, et al. (2018). Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Appl Environ Microbiol 84.

Panek M, Čipčić Paljetak H, Barešić A, et al. (2018). Methodology challenges in studying human gut microbiota—Effects of collection, storage, DNA extraction and next generation sequencing technologies. Sci Rep 8, 5143.

Pasolli E, Asnicar F, Manara S, 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.e20.

Pasolli E, Truong DT, Malik F, et al. (2016). Machine learning meta-analysis of large metagenomic datasets: Tools and biological insights. PLOS Comput Biol 12, e1004977.

Plaza Oñate F, Le Chatelier E, Almeida M, et al. (2018). MSPminer: Abundance-based reconstitution of microbial pan-genomes from shotgun metagenomic data. Bioinformatics 35, 1544–1552.

Quince C, Walker AW, Simpson JT, et al. (2017). Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 35, 833–844.

Raknes IA, and Bongo LA. (2018). META-pipe authorization service. F1000Research 7, ELIXIR-32.

Rampelli S, Soverini M, Turroni S, et al. (2016). ViromeScan: A new tool for metagenomic viral community profiling. BMC Genomics 17, 165.

Ranjan R, Rani A, Metwally A, et al. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun 469, 967–977.

Reyes A, Wu M, McNulty NP, et al. (2013). Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. Proc Natl Acad Sci U S A 110, 20236–20241.

Roux S, Adriaenssens EM, Dutilh BE, et al. (2018). Minimum information about an uncultivated virus genome (MIUViG). Nat Biotechnol [Epub ahead of print]; DOI: 10.1038/nbt.4306.

Rowe WPM, and Winn MD. (2018). Indexed variation graphs for efficient and accurate resistome profiling. Bioinformatics 34, 3601–3608.

Salter SJ, Cox MJ, Turek EM, et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12, 87.

Santana-Pereira ALR, and Liles MR. (2017). Challenges and opportunities in discovery of secondary metabolites using a functional metagenomic approach. In: *Functional Metagenomics: Tools and Applications*. Charles TC, Liles MR, and Sessitsch A, eds. Cham: Springer International Publishing, 119–138.

Savage JH, Lee-Sarwar KA, Sordillo J, et al. (2018). A prospective microbiome-wide association study of food sensitization and food allergy in early childhood. Allergy 73, 145–152.

Schloss PD, and Handelsman J. (2005). Metagenomics for studying unculturable microorganisms: Cutting the Gordian knot. Genome Biol 6, 229.

Schmedes SE, Woerner AE, and Budowle B. (2017). Forensic human identification using skin microbiomes. Appl Environ Microbiol 83, pii: e01672-17.

Schulz F, Alteio L, Goudeau D, et al. (2018). Hidden diversity of soil giant viruses. Nat Commun 9, 4881.

Segata N, Waldron L, Ballarini A, et al. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 9, 811–814.

Sekirov I, Russell SL, Antunes LCM, et al. (2010). Gut microbiota in health and disease. Physiol Rev 90, 859–904.

Simmonds P, Adams MJ, Benkő M, et al. (2017). Consensus statement: Virus taxonomy in the age of metagenomics. Nat Rev Microbiol 15, 161–168.

Sinha R, Abnet CC, White O, et al. (2015). The microbiome quality control project: Baseline study design and future directions. Genome Biol 16, 276.

Stepanauskas R, Fergusson EA, Brown J, et al. (2017). Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nat Commun 8, 84.

Stepanauskas R, and Sieracki ME. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc Natl Acad Sci U S A 104, 9052–9057.

Sutton TDS, Clooney AG, Ryan FJ, et al. (2019). Choice of assembly software has a critical impact on virome characterisation. Microbiome 7, 12.

Taylor CF, Field D, Sansone S-A, et al. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. Nat Biotechnol 26, 889–896.

Taylor CF, Paton NW, Lilley KS, et al. (2007). The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol 25, 887–893.

Thompson LR, Sanders JG, McDonald D, et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551, 457–463.

Tremblay J, Singh K, Fern A, et al. (2015). Primer and platform effects on 16S rRNA tag sequencing. Front Microbiol 6, 771.

Tringe SG, von Mering C, Kobayashi A, et al. (2005). Comparative metagenomics of microbial communities. Science 308, 554–557.

Trompette A, Gollwitzer ES, Yadava K, et al. (2014). Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. Nat Med 20, 159–166.

Truong DT, Tett A, Pasolli E, et al. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res 27, 626–638.

Uritskiy G, and DiRuggiero J. (2019). Applying genome-resolved metagenomics to deconvolute the halophilic microbiome. Genes 10, 220.

van Schaik W. (2015). The human gut resistome. Philos Trans R Soc B Biol Sci 370, 20140087.

Vervier K, Mahé P, and Vert J-P. (2018). MetaVW: Large-scale machine learning for metagenomics sequence classification. Methods Mol Biol Clifton NJ 1807, 9–20.

Wagner J, Paulson JN, Wang X, et al. (2016). Privacy-preserving microbiome analysis using secure computation. Bioinformatics 32, 1873–1879.

Wang Q, Fish JA, Gilman M, et al. (2015). Xander: Employing a novel method for efficient gene-targeted metagenomic assembly. Microbiome 3, 32.

Wang Z, Wang Y, Fuhrman JA, et al. (2019). Assessment of ◀AU10 metagenomic assemblers based on hybrid reads of real and simulated metagenomic sequences. Brief Bioinform.

Wesolowska-Andersen A, Bahl MI, Carvalho V, et al. (2014). Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. Microbiome 2, 19.

Wilson MR, Naccache SN, Samayoa E, et al. (2014). Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med 370, 2408–2417.

Xia Y, Sun J, and Chen D-G. (2018). Bioinformatic analysis of microbiome data. In: *Statistical Analysis of Microbiome Data with R*. Xia Y, Sun J, and Chen D-J, eds. Singapore: Springer Singapore, 1–27.

Yang Y, Jiang X, Chai B, et al. (2016). ARGs-OAP: Online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database. Bioinformatics 32, 2346–2351.

Address correspondence to:
Claudia Carissimi, PhD
Department of Molecular Medicine
''Sapienza'' University of Rome
Vle Regina Elena, 324
Rome 00161
Italy

E-mail: claudia.carissimi@uniroma1.it

Abbreviations Used

NGS = next-generation sequencing
rDNA = ribosomal DNA

AUTHOR QUERY FOR OMI-2019-0073-VER9-LAUDADIO_1P

AU1: Kindly check the edit made to the article heading.

AU2: Kindly check the edit made to the article title.

AU3: Please identify (highlight or circle) all authors' surnames for accurate indexing citations.

AU4: The Publisher requests for readability that all paragraphs be limited to 15 typeset lines. Please adjust all paragraphs accordingly.

AU5: References Bork et al. 2015, Rusch et al. 2007, and Yooseph et al. 2007 are cited in the text, but not found in the reference list. Please check.

AU6: Reference D'Costa et al., 2006 is cited in the text, but not found in the reference list. Please check.

AU7: In Ref. ''Chiu CY 2019,'' please mention the volume number and inclusive page range.

AU8: In Ref. ''LaPierre N 2019,'' please mention the volume number and inclusive page range.

AU9: In Ref. ''Orellana LH 2018,'' please mention the inclusive page range.

AU10: In Ref. ''Wang Z 2019,'' please mention the inclusive page range.

AU11: Please cite Table 1 in the text.