

A Primer on Metagenomics

A Primer on Metagenomics

By

Renu Vyas, Aruna Sivaram, Nayana Patil
and Shama Mujawar

**Cambridge
Scholars
Publishing**



A Primer on Metagenomics

By Renu Vyas, Aruna Sivaram, Nayana Patil and Shama Mujawar

This book first published 2026

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2026 by Renu Vyas, Aruna Sivaram, Nayana Patil
and Shama Mujawar

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 978-1-0364-6425-7

ISBN (Ebook): 978-1-0364-6426-4

TABLE OF CONTENTS

Foreword	xi
Dr Uta Griesenbach	
Chapter 1	1
Understanding Microbiome System: Why it Matters	
Abstract	
Introduction	
1.1 Microbiome in our everyday life	
1.1.1 Food and nutrition	
1.1.2 Health	
1.1.3 Environment	
1.1.4 Energy	
1.2 Need for Metagenomics	
1.2.1 Why microbiology is not sufficient?	
1.2.2 Why is genomics not sufficient?	
1.3 Conclusion	
Exercise: KWHL Chart	
Sample questions	
References	
Chapter 2	23
Metagenomics at Bench: From Sampling to Sequencing	
Introduction	
2.1 Types of Metagenomic analysis	
2.1.1 Bacterial Metagenomics	
2.1.2 Viral Metagenomics	
2.1.3 Clinical Metagenomics	
2.1.4 Cancer Metagenomics	
2.1.5 Metatranscriptomics	
2.2 Preparation of genomic DNA	
2.2.1 Disruption of cell membrane/Cell wall	
2.2.2 Separation and Precipitation of DNA	
2.3 Next generation sequencing	
2.3.1 Library preparation: Fragmentation	
2.3.2 Library preparation: Cloning	
2.3. Library preparation: Sequencing	

2.4 Sequencing technologies	
2.4.1 454 pyrosequencing	
2.4.2 Ion Torrent Technology	
2.4.3 Illumina technology	
2.4.4 Oxford Nanopore technology	
Tutorials	
Questions	
References	
Chapter 3	40
Metagenomics Data Analysis: Reads to Results	
Abstract	
Introduction	
3.1 Filtering data and Assembly	
3.2 Data Prediction and Integration	
3.2.1 Annotation	
3.2.2 FragGeneScan	
3.2.3 MetaGeneMark-2	
3.2.4 MetaGeneAnnotator (MGA)/Metagene	
3.2.5 Orphelia	
3.3 Comparative Metagenomics	
3.4 Cloud Computing and AI	
3.5 Challenges in Data analysis	
3.6 Future trends Metagenome Analysis	
3.7 Conclusion	
Tutorial	
Sample Questions	
References	
Chapter 4	57
Essential Tools and Databases in Metagenomics	
Abstract	
Introduction to Metagenomics	
4.1 Types of Metagenomics	
4.1.1 Metaproteomics	
4.1.2 Meta-transcriptomics	
4.1.3 Metabolomics	
4.1.4 Fluxomics	
4.1.5 Functional metagenomics	
4.1.6 Structural Metagenomics	
4.1.7 Quality Control	
4.2 Importance of Metagenomics	

4.3 Tools in Metagenomics	
4.4 Tools in Metagenomics	
Challenges in Metagenomics	
Emerging Technologies in Metagenomics	
Future Scope of Metagenomics	
Conclusion	
Tutorial	
Q&A	
References	
Chapter 5	145
Clinical Metagenomics: An Introduction	
Abstract	
Introduction	
5.1 Genome-based approaches in medical metagenomics	
5.1.1 Whole genome sequencing	
5.1.2 Targeted next generation sequencing	
5.1.3 Untargeted metagenomic next generation sequencing	
5.2 Metagenomic pipeline for clinical samples	
5.2.1 Sequencing of clinical samples	
5.2.2 Sequence analysis	
5.3 Clinical metagenomics and infectious diseases	
5.3.1 Disease diagnosis and pathogen identification	
5.3.2 Microbial surveillance	
5.3.3 Detection of antibiotic resistance genes	
5.4 Metagenomics in cancer	
5.4.1 Breast cancer	
5.4.2 Gastric cancer	
5.4.3 Cervical cancer	
5.4.4 Hepatocellular carcinoma	
5.4.5 Merkel cell carcinoma	
5.4.6 Nasopharyngeal carcinoma	
5.5 Challenges and limitations of clinical metagenomics	
5.6 Conclusion	
Tutorial	
Question	
References	

Chapter 6	170
Environmental and Ecological Metagenomics	
Abstract	
Introduction	
6.1 Soil Metagenomics	
6.1.1. Application of soil metagenomics in precision agriculture	
6.1.2. Application of soil metagenomics in ionomics	
6.2 Metagenomics and plant growth	
6.2.1 Disease prevention	
6.2.2 Disease diagnosis	
6.2.3 Stress tolerance	
6.3. Aquatic microbiome and metagenomics	
6.3.1 Marine microbiome	
6.3.2 Freshwater microbiome	
6.3.3 Microbiome in drinking water	
6.3.4 Application of metagenomics in marine and freshwater ecosystem	
6.4 Air microbiome	
6.4.1 Application of cloud borne microbial metagenomics and climate change	
6.4.2 Detection of air-borne pathogens	
6.5 Conclusion	
Tutorial	
Sample Questions	
References	
Chapter 7	205
Metagenomics in Bioenergy	
Abstract	
Introduction	
7.1 Metagenomics: Unlocking Uncultured Microbial Diversity	
7.1.1 Key Microbial Communities in Biomass Degradation	
7.1.2 Extremophiles and Their Potential in Bioenergy	
7.1.3 Discovery of Superior Biomass-Degrading Enzymes	
7.2 Microbial engineering and metagenomics	
7.2.1 Engineering Microbial Consortia for Biofuel Production	
7.2.2 Pathway Engineering using Genetic and Synthetic Biology approach	
7.3 Conclusion	
Tutorial	
Questions	
References	

Chapter 8	230
Application of Metagenomics Forensic Sciences	
Abstract	
Introduction	
8.1 Sample processing for forensic metagenomics	
8.1.1 Sample collection	
8.1.2 Storage and transport of samples	
8.1.3 Nucleic acid extraction and Sequencing	
8.2 Microbial communities as forensic evidence	
8.2.1 Microbial signatures	
8.2.2 Microbiome for personal identification	
8.3 Application of forensic metagenomics	
8.3.1 Biological warfare and national security	
8.3.2 Death related investigation	
8.3.3 Geolocation of the crime	
8.3.4 Identifying suspects	
8.4 Advantages of forensic metagenomics	
8.5 Limitations of forensic metagenomics	
8.6 Conclusion	
Tutorial	
Sample Questions	
References	
Chapter 9	248
Metagenomics in Drug Discovery Integrating Omics Data	
Abstract	
9.1 Introduction	
9.1.1 Metagenomics	
9.1.2 Bridging the gap between metagenomics and drug discovery	
9.2 Metagenomics in Drug Discovery	
9.2.1 Influence of metagenomics in drug discovery	
9.2.2 Overview of existing sequencing metagenomics technologies	
9.2.3 Applications in drug discovery	
9.3 Challenges	
9.4 Future scope	
Conclusion	
Q&A	
References	

Chapter 10	266
Ethical Concerns in Metagenomic Studies	
Abstract	
Introduction	
10.1 Impact on Public community	
10.1.1 Ethical concerns in environmental sampling	
10.1.2 Ethical concerns in phylogenetic analysis	
10.2 Impact on individual	
10.2.1 Informed consent	
10.2.2 Concerns regarding privacy	
10.3 Conclusion	
Questions	
References	

FOREWORD

DR UTA GRIESENBACH
IMPERIAL COLLEGE LONDON

Metagenomics is the study of genetic material (typically from microbes) recovered from environmental samples such as skin, soil and water and involves sequencing and aligning of DNA fragments to identify and characterise genetic information” (<https://www.genome.gov/genetics-glossary/Metagenomics>). Importantly, metagenomic analysis does not require culturing of organisms which frequently impacts on accuracy of data generation. To date, metagenomic analysis is used in a large range of disciplines including healthcare, agriculture, bioenergy and biodefence.

This book provides state of the art information related to methods and techniques used, as well as applications and limitations of the technology. Importantly the book serves as a comprehensive guide to the rapidly evolving field of metagenomics. The insights within the chapters will deepen your understanding of the microbial world and its profound impact on life as we know it. The insights within the chapters will deepen your understanding of the microbial world and its profound impact on life as we know it. This book serves as a comprehensive guide to the rapidly evolving field of metagenomics and importantly, is written for a non-expert audience which allows the growing field of metagenomics to become accessible to non-expert readers.

CHAPTER 1

UNDERSTANDING MICROBIOME SYSTEM: WHY IT MATTERS

Abstract

The United Nations Sustainable Development Goals (SDGs) is a blueprint aimed at achieving global welfare. As we move towards the deadline to meet the SDGs, it is pertinent to research on novel strategies, and emphasis is being laid on the role of microbiomes. Characterizing the vast diversity of hundreds to thousands of microbial strains remains an insurmountable challenge, underscoring the limitations of culture-based approaches in comprehensive microbiome profiling. Advances in culture-independent non-traditional approach and sequencing technologies have significantly enhanced our understanding of microbial diversity and its global distribution, enabling a broader perspective on microbial ecology that traditional methods cannot provide. One such approach is the metagenomics which has emerged as a robust tool for comprehensive exploration of microbial population, irrespective of the source of the samples. Compared to conventional microbiological and genomic techniques, metagenomics gives more functional insights and helps in characterization of the yet uncharacterized organisms.

Introduction

Microbes, though microscopic, are an integral part of every process in the biosphere and play an indispensable role in the functioning of the world around them (Larsen *et al.*, 2017). They can thrive in virtually all habitats on Earth due to their ability to adapt swiftly and flexibly to environmental change. Their exceptional potential to drive chemical processes make them central to driving Earth's major biogeochemical cycles, influence human health, and make them indispensable for various industries (Starr *et al.*, 2013).

Microorganisms which include a diverse range of bacteria, archaea, fungi, protozoa, algae, and viruses comprehensively account for about 70 gigatons of carbon constituting $\approx 15\%$ of the global biomass. A recent study estimated that there are between 10^{11} and 10^{12} microbial species inhabiting the earth. Approximately 10^4 microbial species have been successfully cultured in laboratory settings and fewer than 10^5 species are represented by classified genetic sequences (Bar-On *et al.*, 2018; Vitorino *et al.*, 2018). The Earth Microbiome Project (EMP) has cataloged less than 10^7 microbial species which merely uncover the total microbial biodiversity. Not only the number but different types of microbes are estimated to be around $\sim 10^{12}$ microbial species on earth (Thompson *et al.*, 2017). These microbes do not exist as independent species but interact with each other forming a complex microbial interactome so vast and extensive that it is impossible for humans to fully comprehend or perceive (Louca *et al.*, 2019).

1.1 Microbiome in our everyday life

The global microbiome and the interactions among microbial communities are crucial for maintaining the delicate balance between various ecosystems that support all life forms. Microbes are nature's allies in providing eco-friendly solutions for energy production, environmental cleanup, and sustainable agriculture (Thompson *et al.*, 2017; Gilbert *et al.*, 2014). Their ability to transform waste, support agriculture, reduce pollution, and drive green innovations merge seamlessly with almost all of the Sustainable Development Goals (SDGs) established by the United Nations in 2015 as part of the 2030 Agenda for Sustainable Development (Fagunwa *et al.*, 2020). The following section explores how microbial advancements hold transformative potential for achieving these global SDG goals. Microbiomes play a very important role in our everyday life.

1.1.1 Food and nutrition

Holobiont or Metaorganism concept emphasizes the importance of viewing an animal or plants and their associated microorganisms that live in or around the organism as a single functional unit. The interdependent relationship between the host and its microbiome, provide essential nutrients, protect against pathogens, and promote growth of both the partners through symbiotic association (Gamage *et al.*, 2024).

Rhizosphere is one of the complex habitats which is swarmed by incredibly diverse microbial communities, which interact in intricate ways with plants,

soil and neighboring microbial communities. Rhizospheric microorganisms plays an indispensable role in plant growth for example nitrogen-fixing bacteria transform atmospheric nitrogen into ammonia, certain microbes decompose organic matter to release nutrients back into the soil, some convert essential elements such as phosphorus, zinc, potassium etc. to bioavailable forms, while some produce plant hormones like auxins, gibberellins, and cytokinins promoting plant growth on the other hand few microbes are involved in suppressing disease-causing pathogens by producing antibiotics, enzymes, and other metabolites (Mendes *et al.*, 2013).

Accurate identification of microbiome features and probiotics that impact plant health is pivotal for sustainable agriculture. Understanding and harnessing the healthy soil microbiome such as arbuscular mycorrhizal fungi and beneficial bacteria can lead to innovative strategies to enhance crop yields assuring food security and access to nutrition while addressing global issues of poverty and hunger: Goal 2 (Timmis *et al.*, 2017). Similarly, microbial protein offers an eco-friendly alternative to traditional animal protein sources, which can deal with nutritional needs and growing nutritional challenges while minimizing ecological footprints: Goal 15 (Matassa *et al.*, 2016). Furthermore, microbes can be engineered to produce essential nutrients, such as vitamins and amino acids, contributing to fortified foods assisting in managing malnutrition worldwide. Microbes are also explored as biological control agents to manage pests and diseases in crops, reducing the need for harmful synthetic chemicals and promoting eco-friendly agricultural practices (Goal 6) (Crowther *et al.*, 2024).

Microbial antagonists inhibit harmful pathogens, while endophytes promote healthier plants and reduced crop losses without the use of synthetic fertilizers and pesticides. Likewise microbial inoculants are known to boost plant immunity and improve resistance to diseases further reducing the need of synthetic agrochemicals. By reducing chemical inputs and improving resource efficiency, microbes can help mitigate the bioaccumulation of harmful chemicals in the food chain and environment allowing us to achieve Goal 6 and Goal 14 (Fagunwa *et al.*, 2020; Crowther *et al.*, 2024). By leveraging the potential of microbes to break down organic waste into valuable products like biocompost and biogas, contributing to a circular economy can help us achieve Goal 8 (Timmis *et al.*, 2017; Singh). In addition, microbes can also be involved in food safety by employing them in detecting pathogens and contaminants. Microbial assays can be used to monitor food quality, ensuring consumer health through microbial assays that monitor food quality (Goal 2) (Carlino *et al.*, 2024).

1.1.2 Health

Human Microbiome refers to the human body as a niche populated by trillions of microbes in a microbial-to-human cell ratio of approximately 1:1. Majority of these microorganisms belong to *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria*, *Fusobacteria*, and *Verrubomicrobia* phylum and form a symbiotic relationship with the human host. *Firmicutes* and *Bacteroidetes* phyla constitute about 90% of the gut microbiota and the *Firmicutes/Bacteroidetes* (F/B) ratio plays a crucial role in maintaining gut homeostasis. For example, an increased F/B ratio is associated with obesity while decreased ratio is associated with inflammatory bowel disease. The imbalance of gut flora leads to increase in the abundance of pathobionts and has been linked to digestive disorders, metabolic syndrome like diabetes, autoimmune diseases, cardiovascular disorders, cancer, hypertension and mental health conditions such as depression and anxiety (Martino *et al.*, 2022; Afzaal *et al.*, 2022).

Around 500 to 1,000 species of microbes are estimated to live in the human gut and are often referred to as the “second gene pool.” They are involved in digestion of food, fermenting dietary fibres, synthesis of vitamins and amino acids, boosting immunity and shielding us from pathogens. The trillions of microbes forming the gut microbiome is also considered an endocrine gland because it produces and secretes bioactive compounds that influence the body’s hormonal balance. These include neurotransmitters like serotonin, which affects mood and behavior, and other signaling molecules that can impact insulin sensitivity, metabolic rate, and even stress response (Ahlman *et al.*, 2001).

The understanding of the human microbiome has led to the inception of a new domain termed Microbiome medicine making notable strides towards SDGs Goal 3 of good health and wellbeing. Microbiome medicine encompasses emerging applications such as faecal microbiota transplantation, personalized microbiome-based prebiotic and probiotic therapies. Microbiome-based diagnosis, microbial biomarkers, synthesis of vaccines, next-generation symbiotic, antibiotics and integrating microbiome interventions with immunotherapies are some of the latest innovative microbiome based health interventions (O’Toole *et al.*, 2023). If these innovations can be made affordable and lend them to use in low income countries, the cost of accessing them can contribute to SDG Goal 10 (Fagunwa *et al.*, 2020; Crowther *et al.*, 2024).

Microbial bioremediation reduces pollutants and toxins in the environment, indirectly preventing diseases caused by contaminated water, air, and soil, (Goal 6) thereby improving public health (Kuppan *et al.*, 2024). Additionally, the use of microbes to combat disease transmission in insect vectors has proven successful; for instance, the bacterium *Wolbachia* and *Bacillus thuringiensis israelensis* have been employed to infect *Aedes aegypti* mosquitoes, decreasing their ability to spread dengue, Zika, and chikungunya viruses. Similarly, *Serratia marcescens* in sandflies blocks the transmission of *Leishmania*, showcasing the potential of microbial interventions in disease control and environmental health (Katake *et al.*, 2023).

1.1.3 Environment

The environmental microbiome is a cornerstone of Earth's ecosystems, in the previous section we have already discussed the role of soil microbiome and its symbiotic relationships with plants. Along with this the soil microbes modulate and maintain the carbon cycle in the atmosphere through photosynthesis (Gougoulias *et al.*, 2014). The oceanic microbes such as phytoplankton, Diatoms, Coccolithophores, Cyanobacteria fix approximately 40-50 gigatons of carbon per year accounting to 50% of the global carbon dioxide (CO₂) through photosynthesis. When these microbes become part of marine food webs and eventually die their carbon-rich remains sink to the ocean floor (Pedros-Alio 2006). This transformation plays an important role in global biogeochemical cycles, composition of Earth's atmosphere, and on global climatic conditions thereby contributing towards sustaining the planet's habitability. Microbes can influence the fate of carbon dioxide in the atmosphere, as they remove CO₂ from the atmosphere, thereby mitigating some of the anthropogenic release of CO₂ and other greenhouse gases such as methane (CH₄), and nitrous oxide (N₂O) (Ma *et al.*, 2023; Van Den Berghe *et al.*, 2024).

Moreover we depend on microbes to remediate toxins in the environment. Successful bioremediation has been implemented for a wide range of contaminants, including heavy metals, petroleum hydrocarbons, pesticides, and plastics, making it a versatile approach for environmental restoration. Both naturally occurring and genetically engineered microbes can be employed in bioremediation efforts (Pant *et al.*, 2021). Advancing bioremediation strategies and enhancing their application in environmental cleanup efforts, ultimately contributing to a healthier planet. Ongoing research focused on improving bioremediation techniques, development of microbial consortia and the use of advanced biotechnological methods to

enhance pollutant degradation aligns with the Sustainable Development Goals (SDGs), particularly SDG 15: Life on Land, SDG 13: Climate Action, SDG 6: Clean Water and Sanitation and SDG 14: Life Below Water (Carlino *et al.*, 2024; Fagunwa *et al.*, 2020; Crowther *et al.*, 2024).

Bioremediation techniques help in restoring contaminated soils, remove toxins, support biodiversity, and contribute to ecosystem health thereby promoting sustainable use of terrestrial ecosystems and combating biodiversity loss. Through microbial bioremediation of polluted water bodies we can improve water quality and access ensuring availability and sustainable management of water and sanitation for all. Mitigation of marine pollution due to oil spills, plastic and heavy metals via bioremediation will lead to protecting aquatic life and maintaining marine biodiversity in doing so conserving the marine ecosystem (Aliko *et al.*, 2022). By employing microbial carbon sequestration, methane capture and developing biofuels from microbial biomass can reduce the greenhouse gas emission. Recognizing the role of microbes in climate action is critical for developing effective environmental policies and strategies for fostering resilient ecosystems that can support life on Earth (*et al.*, 2018).

1.1.4 Energy

Microbes are remarkable in their ability to harbor and produce a diverse array of enzymes that facilitate the conversion of several complex organic substrates into simpler molecules essential for metabolism and survival.

Soil and plant microbiomes mainly produce cellulases, hemicellulases, and lignases that coordinate to deconstruct plant biomass, facilitating the conversion of cellulose, hemicellulose, and lignin into usable sugars and smaller compounds (Jansson *et al.*, 2018).

Even if lignocellulose is the third most abundant natural polymer estimated to 120×10^9 tons annually are under-utilized for biofuel production due to their recalcitrant nature. However microbial pre-treatment and enzymatic digestion are used to enhance the lignocellulosic biomass degradation. Pretreatment of lignocellulosic wastes with rumen microbiome was reported to enhance methane yield by 82.6% (Zhang *et al.*, 2016) along with massive reduction of volatile solids. The rumen microbiome consists of fibrolytic microorganisms, methanogenic archaea and cellulose-degrading bacterial consortium (RCBC) contributes towards the major plant cell wall degraders leading to liquid fuel, biohydrogen and biomethane production.

On the other hand microalgal biomass is considered a third-generation biofuel feedstock due to their high lipid content, rapid growth rate, ease of extraction and carbon sequestration capability has revolutionized the renewable energy sector. Biogas and biobutanol are identified as a substitute for natural gas. Anaerobic fermentation by methanogenic bacteria are critical in biogas digesters, breaking down organic waste into methane and carbon dioxide while anaerobic fermentation by *Clostridium* strains produce butanol, acetone and ethanol (Kharitonov *et al.*, 2021). MFC and microbial electrolysis cells (MECs) can transform chemical energy stored in organic materials into bio-electrical energy through microbes. This technology can generate electricity while simultaneously treating wastewater, making it a dual-purpose solution for energy production and waste management (Obileke *et al.*, 2021).

Biofuels from organic materials, including non-food agricultural and municipal waste can significantly reduce greenhouse gas emissions compared to fossil fuels. Algal biomass captures carbon dioxide directly from the atmosphere or industrial waste gases, converting them into fuels thus contributing to net-zero carbon emissions and helping mitigate climate change. Methanotrophic bacteria can utilize methane as a carbon source and help reduce methane emissions, a potent greenhouse gas (van der Ha *et al.*, 2012).

Clean energy production, capturing and utilizing greenhouse emissions, and providing innovative solutions for sustainable energy systems accentuate the importance of microbial biotechnology in addressing global energy challenges and promoting environmental sustainability. Thus microbes can be harnessed for clean energy production, contributing to the achievement of affordable and clean energy (Goal 7), Industry, Innovation and Infrastructure (Goal 9), responsible consumption (Goal 12) and supporting climate action (Goal 13). If the waste generated at the source can be valorised to heat and power will lead to reducing the negative environmental impacts of cities and make them more sustainable (Goal 11) (Diaz-Troya *et al* 2024).

Thus Microbes play a crucial role in achieving the Sustainable Development Goals (SDGs) by supporting agriculture, health, environmental sustainability, and clean energy. Figure 1 maps the diverse applications of microbes which contributes towards the sustainable development and a healthier planet.

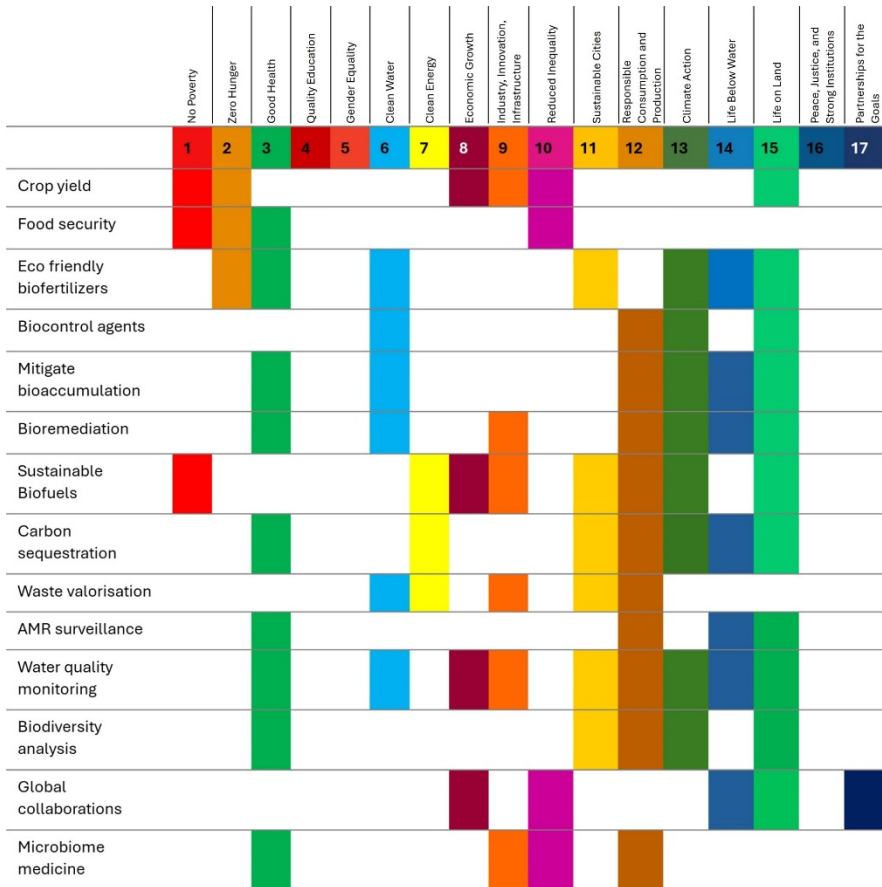


Figure 1. Mapping of metagenomics applications with the United Nations Sustainable Development Goals (SDGs). This figure illustrates how metagenomics contributes to various SDGs. The interconnected roles of metagenomics in achieving global sustainability are highlighted.

1.2 Need for Metagenomics

The conventional reductionist approach in microbiology has primarily emphasized on isolating the microbial species and cultivating them in artificial media under lab conditions that try to mimic but do not reflect their natural ecological environments. Metagenomics has now gained momentum in study of microbes as it helps in characterizing the uncharacterized.

1.2.1 Why is microbiology not sufficient?

A combination of microscopy, biochemical assays and propagating microbes on selective media are undertaken in culture-dependent methods to profile complex microbial ecosystems. A key advantage of these techniques is their ability to identify the viable portion of the microbiome. Nonetheless, because growth of pure culture is heavily influenced and dependent on culturing conditions, environmental factors, and the physical state of the microbial culture, culture-dependent methods result in diagnostic bias. In addition to being time-consuming, microbial culture techniques precludes non-culturable, fastidious and obligate symbionts (Lok *et al.*, 2015).

As per the great plate-count anomaly postulated by Staley and Konopka in 1985 only 0.1 - 1.0% of the living bacteria present in soils and aquatic ecosystems can thrive under laboratory conditions as most of these microbes are resistant to cultivation. The vastness of microbial biodiversity makes it infeasible for pure-culture methods to fully capture or quantify these communities.

The discrepancy between the numbers of microorganisms visible via microscopy and the colonies obtained from laboratory cultivation is several orders of magnitude for most soils, and overcoming the “great plate count anomaly” (Staley and Konopka, 1985) in order to access a greater diversity of bacteria has become one of the most significant challenges and opportunities in the field of molecular microbial ecology.

Intraspecies variability and synergistic interactions within community clusters are challenging to replicate using culture-based methods in most laboratory environments. To address this, modern culture-independent techniques have emerged, focusing on the amplification and sequencing of phylogenetic markers or employing multi-gene metabarcoding combined with advanced bioinformatics analysis. These approaches provide deeper insights into the complexities of microbial communities that traditional methods fail to capture (Forbes *et al.*, 2017).

Culture-Dependent methods primarily focusing on the presence of organisms do not provide detailed information on their genome and gene functions. Gene sequencing has enabled researchers to study the functional genes present in microbial communities and contribution of these genes to their ecosystem (Burke *et al.*, 2011). Even if the first microbial genome, that of *Haemophilus influenzae*, was sequenced from pure cultures in 1995 (Fleischmann *et al.*, 1995), over the past decade, significant advancements

have been made with culture-independent techniques, leading to the sequencing of approximately 1,000 fungal genomes and over 330,000 bacterial and archaeal genomes as of 2024 (Eisenhofer *et al.*, 2024). This breakthrough has been driven by the high-throughput sequencing technologies that have revolutionized our ability to comprehend microbial diversity and genetic profile of an environment in a fraction of the time. Genomics allows for the study of entire microbial communities directly from environmental samples and reveals a more accurate representation of microbial diversity, including unculturable and most importantly previously unknown species.

The discrepancy between the numbers of microorganisms visible via microscopy and the colonies obtained from laboratory cultivation is several orders of magnitude for most soils, and overcoming the “great plate count anomaly” (Staley and Konopka, 1985) in order to access a greater diversity of bacteria has become one of the most significant challenges and opportunities in the field of molecular microbial ecology. These culture-independent methods have revolutionized our ability to study microbial diversity and function at a scale that pure culture approaches could not achieve. Predominant among them is community ribosomal RNA (rRNA) phylotyping, where polymerase chain reaction (PCR) makes copies of rRNA extracted directly from environmental samples, which are then cloned and sequenced without need of the culturing the organism. Decoding sequences through genomics has deepened our insights; however, similar to traditional pure culture methods, standard genomics has limitations in its capacity to clarify the interactions and dynamics of microbial communities (Wooley *et al.*, 2010). Phylotyping provides credible information about “Who is there?” but due to microdiverse clusters within the niche can only guesstimate “What are they doing?”

As illustrated in Figure 2 each milestone in the evolution of metagenomics led to major breakthroughs, innovations, and advancements over time. The timeline initiates from aseptic handling of microbes to the concept and early research on sequencing DNA, followed by significant technological improvements making it more affordable. The commercial applications and widespread adoption in exploring ecosystems have generated a humongous amount of data (Table 1). Each milestone represents a crucial step in shaping the technology’s progress and impact on society.

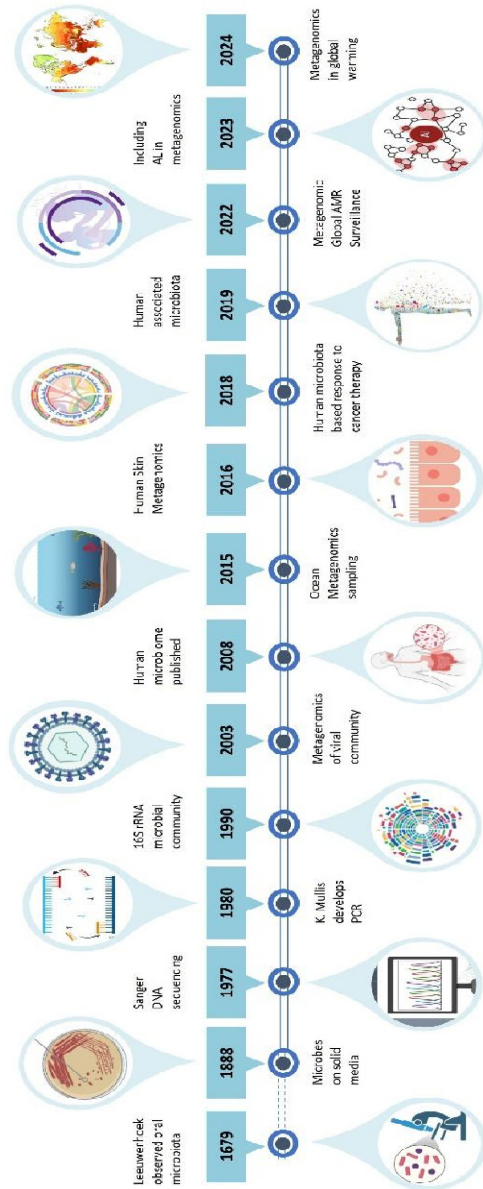


Figure 2. Metagenomic era highlighting the milestones in the development of Metagenomic studies achieved around the globe.

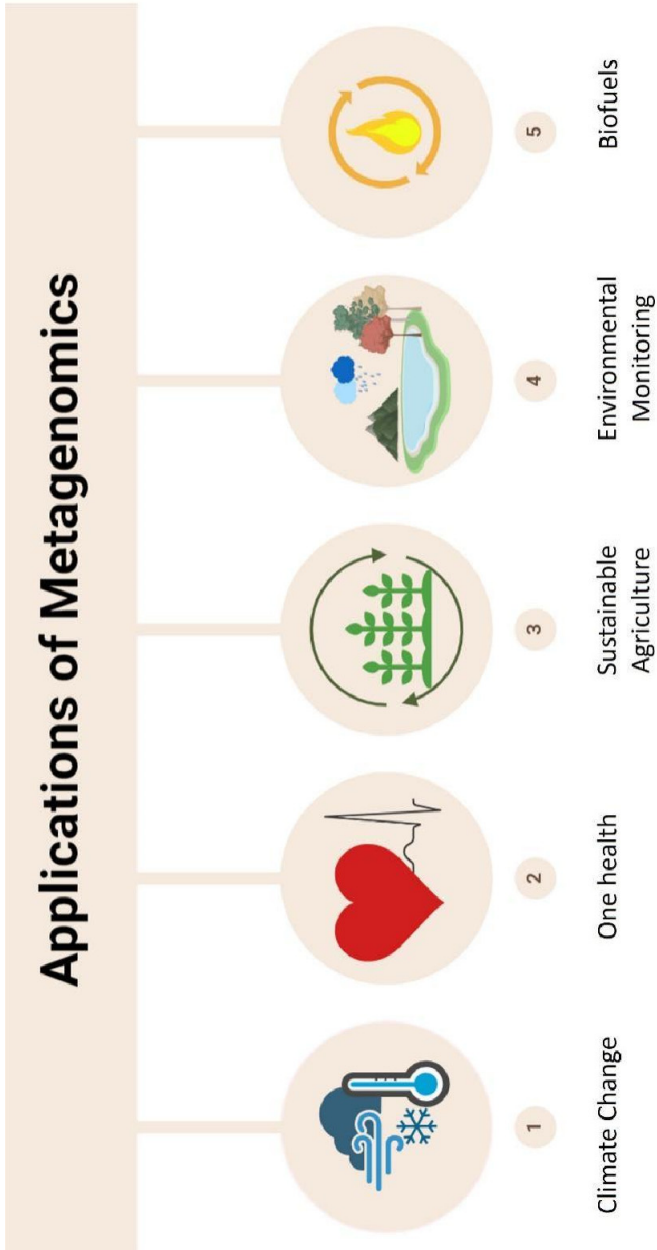
1.2.2 Why Genomics is not enough

It is an arduous task to fully capture the genomic diversity of the microbial world by sequencing a single set of genes or complete genome using traditional genomic approaches. To truly understand microbial diversity, more comprehensive methods are needed that go beyond standard genome sequencing. As one comes to appreciate the true extent of microbial diversity, it becomes clear that understanding complex and collective interactions requires methods that go beyond traditional pure-culture and single-whole-genome approaches (Sharon *et al.*, 2013). If one wants to overcome the biases associated with culture dependent methods and address function driven questions then sequencing alone is inadequate on its own. It has to be supplemented with metagenomics, which can analyze mixed microbial populations and reveal insights into both the presence and function of microorganisms within their ecosystems (Quince *et al.*, 2017).

The top-down approach of metagenomics (Amit G *et al.*, 2023) focuses on the entire genomes and their behavior as metabolic consortia answering not only the “What kinds of microbes are present?”, “What metabolic activities are occurring”, “How are microorganisms linked and interacting?” questions, but also answers, “How does the community function collectively contributing to its ecosystem”.

The term metagenomics was first used by Jo Handelsman, Jon Clardy, Robert Goodman, Sean Brady and first appeared in 1998 publication. Metagenomics can be defined as a cultivation-independent analysis of the genetic material of microbial communities, enabling the study of genes and functions at a community level. It aims to understand the interactions and behaviors of microbes and their role in the biosphere at the genomic scale. This community metagenomics maximizes our understanding of the genetic composition of communities and how these genes work in synchronization that contribute to the functioning of the ecosystem. The lens of metagenomics

Figure 3 (next page). Applications of metagenomics across various fields. (1) Climate change research—studying microbial responses to environmental shifts. (2) Human health—identifying disease-related microbiomes and novel therapeutics. (3) Agriculture—enhancing soil health and plant productivity through microbial interactions. (4) Environmental monitoring—assessing biodiversity and detecting pollutants in ecosystems. (5) Bioenergy—exploring microbial pathways for sustainable energy production. These applications align with global sustainability and scientific advancements.



helps to decipher the complex interactions like gene exchange, metabolic interplay, cell complementarity within and between microbiome clusters that dictate their response to physical and biological variables (Berg *et al.*, 2020).

As summarised in Figure 3, Metagenomics can contribute to advancements in many fields. Some of them are summarized below.

- Climate change: Lithosphere and hydrosphere metagenomics help in understanding of Earth's biogeochemical cycles and greenhouse gas emissions, providing insights into how microbes can be harnessed to reduce carbon dioxide levels and mitigate global warming thereby restoring environmental health (Ciais *et al.*, 2014).
- Metagenomics identifies novel enzymes from microbial communities capable of breaking down lignocellulose, aiding in biofuel production. Understanding algal microbiomes and its lipid content improves its biochemical transformation into biodiesel. Sequencing methanogenic microbial communities enhances the production of biogas along with reducing greenhouse gas emissions (Xing *et al.*, 2012; Campanaro *et al.*, 2016).
- Metagenomics enhances the effectiveness of bioremediation processes by uncovering microbial genes involved in degrading hydrocarbons, pesticides, and heavy metals, enabling cleanup of contaminated environments and sensitive monitoring of remediation (Bharagava *et al.*, 2019).
- Sequencing soil microbiomes identifies beneficial microbes assisting in developing microbial consortia to boost crop yields sustainably. Metagenomics finds microbial agents for biopesticides, reducing reliance on chemical pesticides (Gamage *et al.*, 2024).
- Metagenomics rapidly identifies potential emerging pathogens, bioterror agents, and helps in resilience planning against unknown microbial threats in real-time (Oliveira *et al.*, 2020).
- Metagenomics identifies novel enzymes and unique metabolites from uncultivable microbes which can have application in healthcare, food and environmental sectors (Milshteyn *et al.*, 2014).

Table 1. List of Super studies in metagenomics that analyze massive datasets of microbial communities across multiple environments.

Title	Description
Tara Oceans	The Tara Oceans expedition has collected seawater samples from all over the globe, offering in-depth insight into microbial life in the world's oceans.
Earth Microbiome Project	The Earth Microbiome Project is a wide ranging collaborative project includes 96 different studies, comprising approximately 26k individual sequencing runs from a diverse range of biomes.
NASA GeneLab Microbiome (MANGO)	Understand how microbial communities adapt to spaceflight and related terrestrial stresses Project MANGO provides access to the microbiomes of the international space station and a range of environments selected by NASA GeneLab and the JPL interplanetary protection team.
HoloFood	Holistic approach to improve the efficiency of food production systems by deciphering the biomolecular and physiological processes triggered by incorporating feed additives and novel sustainable feeds in farmed animals. For more information, see the project site.
Malaspina	The Malaspina circumnavigation expedition was an interdisciplinary research project undertaken in 2010-2011 to assess the impact of global change on the oceans and explore their biodiversity.
AtlantECO	The EU-funded AtlantECO project aims to develop and apply a novel, unifying framework that provides knowledge-based resources for a better understanding and management of the Atlantic Ocean and its ecosystem services.

FindingPheno	FindingPheno is creating an integrated computational framework for hologenomic big data, providing the tools to better understand how host-microbiome interactions can affect growth and other outcomes.
National Mouse Genetics Network (NMGN) Microbiome Cluster	The Microbiome Cluster of the National Mouse Genetics Network will achieve two complementary goals: it will develop an experimental pipeline for creating and studying mouse models of human genetic diseases involving barrier surfaces,

1.3 Conclusion

Conventional microbiological methods have highlighted the immense potential of microbes in various domains. However, the advent of genomics and metagenomics marks a paradigm shift, offering a fresh perspective on unprecedented ways to study the microbial world. These advanced approaches promise not only to transform the field of microbiology but also reveal the dynamic interactions within entire communities representing ecological niche to human health and beyond. While microbiology has long been acquainted with community ecology, the integration of genomics has created hitherto unheard-of opportunities for research. Metagenomics, in particular, has unveiled the immense genetic and metabolic diversity within microbial communities, illuminating their remarkable capabilities and indispensable roles in ecosystems.

Microbial communities thrive in virtually every environment, yet their composition and ecological roles remain largely unexplored, leaving millions of potential metagenomics studies waiting to be conducted. Metagenomics has begun to uncover these hidden worlds, offering astonishing insights with vast practical applications. These groundbreaking revelations and their implications will be explored in greater detail in the following sections of this book.

KWHL Chart

- **K – What I Know**
- **W – What I Want to Know**
- **H – How I Will Learn**
- **L – What I Learned**

K (What I Know)	W (What I Want to Know)	H (How I Will Learn)	L (What I Learned)
Microbiomes are essential for health, environment, and industry.	How do metagenomics help overcome the limitations of traditional microbiology?	Reading academic papers, textbooks, and research articles on microbiome studies.	
Metagenomics is a powerful tool for studying microbial diversity.	How does metagenomics contribute to sustainable development goals (SDGs)?	Attending lectures, webinars, and expert talks on metagenomics and its applications.	
Microbes influence food, energy, and disease.	What are the major applications of metagenomics in bioremediation and energy production?	Conducting experiments, analyzing case studies, and reviewing datasets.	
Traditional microbiology cannot fully capture microbial diversity.	What are the latest advancements in sequencing technologies for metagenomics?	Engaging in discussions with experts and participating in laboratory work.	

Sample questions

1. What is the significance of microbes in the Earth’s biogeochemical cycles and their role in the environment?
2. How does the rhizosphere microbiome contribute to plant health and sustainable agriculture?
3. What is the Firmicutes/Bacteroidetes (F/B) ratio, and how does it influence human health?

4. In what ways does microbiome research align with the United Nations Sustainable Development Goals (SDGs)?
5. What are the major limitations of traditional culture-based microbiological methods in studying microbial diversity?
6. How has metagenomics revolutionized our understanding of microbial communities compared to traditional genomics?
7. What are some key microbial-based bioremediation strategies for environmental cleanup?
8. How can microbes be utilized for sustainable energy production, and what are the major biofuel sources?
9. What are some innovative applications of microbiome-based medicine in human health?
10. How does metagenomics aid in identifying novel enzymes and bioactive compounds for various industrial applications?

References

1. Starr, Mortimer P *et al.*, eds. *The prokaryotes: a handbook on habitats, isolation and identification of bacteria*. Springer Science & Business Media, 2013.
2. Bar-On, Yinon M., Rob Phillips, and Ron Milo. "The biomass distribution on Earth." *Proceedings of the National Academy of Sciences* 115.25 (2018): 6506-6511.
3. Vitorino, Luciana Cristina, and Layara Alexandre Bessa. "Microbial diversity: the gap between the estimated and the known." *Diversity* 10.2 (2018): 46.
4. Parks, Donovan H., *et al.* "Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life." *Nature microbiology* 2.11 (2017): 1533-1542.
5. Louca, Stilianos, *et al.* "A census-based estimate of Earth's bacterial and archaeal diversity." *PLoS biology* 17.2 (2019): e3000106.
6. Thompson, Luke R., *et al.* "A communal catalogue reveals Earth's multiscale microbial diversity." *Nature* 551.7681 (2017): 457-463.
7. Gilbert, Jack A., Janet K. Jansson, and Rob Knight. "The Earth Microbiome project: successes and aspirations." *BMC biology* 12 (2014): 1-4.
8. Fagunwa, Omololu E., and Afolake A. Olanbiwoninu. "Accelerating the sustainable development goals through microbiology: some efforts and opportunities." *Access microbiology* 2.5 (2020): e000112.