International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7, 2025
https://theaspd.com/index.php

Microbial Diversity And Environmental Ecology Empowered By Bioinformatics Tools And Applications

Beladel Ahmed[1]*, Boutaiba Saad[2], Beladel Brahim[3]
[1]Laboratory of Exploration and Valorization of Steppe Ecosystems, University of Djelfa, P.O. Box 3117, Djelfa, Algeria, beladel.ahmed@univ-djelfa.dz
[2]Department of Biology, Faculty of Natural and Life Sciences, University of Djelfa, P.O. Box 3117, Djelfa, Algeria.
[3]Department of Physics, Faculty of Science and Technics, University of Djelfa, P.O. Box 3117, Djelfa, Algeria.

Abstract
Bioinformatics has profoundly transformed microbiology by enabling high-throughput analysis and interpretation of complex biological data. From the foundational era of sequence alignment tools to the current landscape of multi-omics integration, machine learning, and quantum-informed approaches, bioinformatics has become indispensable for decoding microbial diversity, community dynamics, ecosystem functions, and pathogen behavior. Its applications now extend far beyond taxonomy and ecology, encompassing systems biology modeling, pathogenomics, structure-based drug docking, toxin gene profiling, and gene therapy design. This review overview traces the chronological evolution of bioinformatics tools and highlights their pivotal roles in both environmental and biomedical microbiology. We examine key technological milestones, showcase representative case studies from diverse and often understudied ecosystems, and assess the strengths and limitations of current methodologies. Emerging frameworks from quantum biology are also being explored to explain complex phenomena such as electron tunneling in enzymes and mutation dynamics. Finally, we explore future directions shaped by real-time sequencing, AI-driven analytics, and quantum computing technologies that promise to redefine our understanding and manipulation of the microbial world for health, biotechnology, and environmental sustainability.
Keywords: *Bioinformatics, Microbial Diversity, Multi-Omics Integration, Machine Learning, Metagenomics, Environmental Microbiology.*

1. INTRODUCTION

The study of microorganisms in their natural habitats has long captivated microbiologists, ecologists, and environmental scientists. Microbial communities play essential roles in ecosystem functioning, biogeochemical cycling, and even human health. However, traditional microbiological methods, which relied heavily on culture-based techniques, overlooked most microorganisms that are not readily cultivable under laboratory conditions. These limitations significantly constrained our understanding of microbial diversity and ecological function (Amann et al., 1995).

A paradigm shift began with the advent of DNA sequencing technologies in the late 20th century, particularly with the development of 16S rRNA gene sequencing as a tool for identifying and classifying microorganisms independent of cultivation (Woese & Fox, 1977). This breakthrough was further accelerated by the emergence of high-throughput next-generation sequencing (NGS) technologies in the 2000s, enabling the generation of massive datasets from diverse environments, including soils, oceans, hypersaline lakes, and human-associated microbiomes (Caporaso et al., 2012).

These advances ushered in the era of bioinformatics a multidisciplinary field integrating biology, computer science, and statistics to manage, process, and interpret complex biological data. Bioinformatics tools have become indispensable for assembling, annotating, visualizing, and interpreting genomic and metagenomic information. Widely used platforms such as QIIME (Caporaso et al., 2010), Mothur (Schloss et al., 2009), and MG-RAST (Keegan et al., 2016) allow researchers to investigate microbial community structures, functional capacities, and ecological interactions at unprecedented resolution and scale.

Moreover, bioinformatics has deepened our understanding of evolutionary relationships, horizontal gene transfer, and the discovery of novel genes and pathways involved in stress adaptation, antimicrobial resistance, and niche specialization. More recently, machine learning algorithms and multi-omics integration have enabled predictive modeling of microbial functions and dynamics, offering high-resolution insights into ecosystem-level interactions.

This paper aims to synthesize the evolution and current state of bioinformatics tools in microbiology, microbial ecology, and biodiversity research. We explore how these tools have expanded our understanding of microbial life particularly in diverse and extreme ecosystems and provide theoretical insights into their broader scientific, ecological, and biomedical implications. Finally, the rise of AI and quantum computing is creating space for new approaches to simulate complex microbial behaviors and biochemical reactions in silico (Pal et al., 2023).

2. Historical Background: Early Bioinformatics Tools in Microbiology

The term *bioinformatics* emerged in the 1970s, initially referring to the application of computational methods for managing and analyzing biological data. Its scientific and public recognition expanded dramatically in the 1990s, catalyzed by large-scale efforts like the Human Genome Project (Luscombe et al., 2001). In microbiology, bioinformatics enabled a pivotal shift from culture-dependent approaches, such as manual colony counting (Prescott et al., 1996), to genome- and sequence-based investigations using early sequence alignment tools like ClustalW (Thompson et al., 1994), ushering in the era of high-throughput, data-driven microbial exploration (see Table 1).Early bioinformatics tools were primarily developed for fundamental analytical tasks (illustrated in Figure A), including sequence alignment, phylogenetic tree construction, and taxonomic classification. These tools introduced computational rigor, reproducibility, and scalability to microbial research:

- BLAST (Basic Local Alignment Search Tool): Developed by Altschul et al. (1990), BLAST utilizes a heuristic algorithm for rapid local alignment of nucleotide or protein sequences against reference databases, efficiently identifying high-scoring segment pairs (HSPs) and enabling fast, sensitive homology detection.
- FASTA: Introduced by Pearson and Lipman (1988), FASTA also performs local sequence alignments using a dynamic programming approach, offering an early but powerful alternative to BLAST for detailed nucleotide and protein similarity analyses.
- PHYLIP (Phylogeny Inference Package): Created by Felsenstein (1989), PHYLIP provides tools for constructing phylogenetic trees through evolutionary algorithms such as maximum parsimony, distance matrix methods, and maximum likelihood estimation.
- MEGA (Molecular Evolutionary Genetics Analysis): First released by Kumar et al. (1994), MEGA facilitates tree-building using methods like neighbor-joining, UPGMA, and maximum likelihood, and remains a widely used platform for evolutionary analysis and visualization.
- RDP (Ribosomal Database Project): The RDP is a curated 16S rRNA sequence repository that supports robust taxonomic classification using a naïve Bayesian classifier with statistical confidence measures (Wang et al., 2007).
- Greengenes: Developed by DeSantis et al. (2006), Greengenes is another 16S rRNA gene database that provides precomputed alignments and taxonomic trees and has been broadly integrated into workflows such as QIIME (Caporaso et al., 2010).

The emergence of these tools marked a turning point in microbiology, enabling researchers to analyze microbial genetic sequences at scale, identify unculturable taxa, trace evolutionary lineages, and characterize complex microbial communities across diverse ecosystems.

Importantly, the conceptual development of these tools was influenced by earlier advancements in mathematics, computer science, and operations research. During World War II, for instance, fields like linear programming and systems engineering played key roles in optimizing military logistics such as the strategic deployment of tank divisions (Winsberg, 2010). These same principles later found fertile ground in genomics, where algorithm design, data modeling, and computational optimization became central to decoding the microbial world.

ERA / TIMEFRAME	TOOL / APPROACH	MAIN FUNCTION	KEY FEATURES	REFERENCES
1980S	Manual colony counting	Quantify culturable microbial populations	Plate-based methods; non-digital	Prescott et al., 1996
1990S	BLAST	Sequence similarity search	Local alignment algorithm	Altschul et al., 1990
EARLY 2000S	ClustalW / MEGA	Multiple sequence alignment, phylogenetic analysis	GUI-based; MSA and trees	Thompson et al., 1994; Tamura et al., 2007
MID-2000S	GitHub	Code sharing, collaboration, reproducibility	Version control, community engagement, open source	Blischak et al., 2016
2005–2010	QIIME (v1), MOTHUR	Microbial 16S data analysis	OTU-based classification, alpha/beta diversity	Caporaso et al., 2010; Schloss et al., 2009
2010–2015	MG-RAST, SILVA, RDP, KEGG	Functional and taxonomic annotation	Curated databases and web servers	Meyer et al., 2008; Quast et al., 2013; Kanehisa et al., 2023
2015–2020	QIIME 2, DADA2, DIAMOND	High-resolution and fast matching	ASV, reproducibility, faster alignments	Bolyen et al., 2019; Callahan et al., 2016
2020–PRESENT	DeepMicrobes,	AI-based genome prediction, pattern recognition	Deep learning; multi-omics integration	Liang et al., 2020; Wang et al., 2022

Table 1: Chronological Evolution of Bioinformatics Tools and Approaches in Microbiology
This table outlines the timeline of key technological milestones in microbial bioinformatics, highlighting the shift from traditional, manual methods (e.g., colony counting in the 1980s) to high-throughput, AI-powered platforms of the modern era. Each era reflects the dominant tools, their primary applications (e.g., sequence alignment, taxonomic classification, genome prediction), and distinguishing features such as algorithmic innovation, GUI interfaces, and open-source development. The emergence of tools like QIIME, MG-RAST, and DeepMicrobes represents the growing integration of metagenomics, multi-omics, and machine learning in microbial research. Together, these developments chart the field's transition into a data-driven, systems-level discipline.

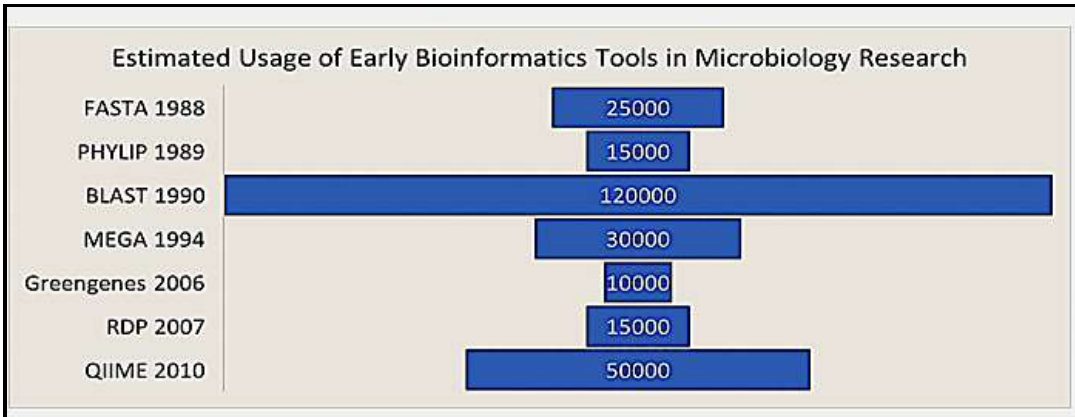

Figure A: This horizontal bar chart presents the approximate number of research papers that have cited or used key bioinformatics tools since their introduction. BLAST (1990) stands out as the most widely used tool, with over 120,000 citations, followed by QIIME (2010) and MEGA (1994). These figures reflect each tool's impact on microbial sequence analysis, phylogenetics, and microbial ecology workflows.

3. Advances in High-Throughput and Metagenomic Tools

The rapid evolution of high-throughput sequencing (HTS) platforms such as Illumina (Illumina Inc., USA), Ion Torrent (Thermo Fisher Scientific, USA), and PacBio (Pacific Biosciences, USA) has revolutionized microbial ecology by enabling the deep exploration of complex microbial communities across diverse environments (Figure B). This technological leap has been accompanied by the development of increasingly sophisticated bioinformatics pipelines and software, built using languages such as Python, R, C++, Perl, and JavaScript, and often deployed through environments like RStudio, Jupyter Notebooks, and Linux-based shells.

3.1 Sequence Processing and Denoising

A key methodological shift has been the transition from operational taxonomic units (OTUs) to amplicon sequence variants (ASVs), which offer improved resolution and reproducibility.

- QIIME 2 (Bolyen et al., 2019), a Python-based pipeline, integrates algorithms such as UCLUST (Edgar, 2010), VSEARCH (Rognes et al., 2016), and DADA2 (Callahan et al., 2016) to model sequencing errors and infer true biological sequences (ASVs).
- Mothur (Schloss et al., 2009), developed in C++ and Fortran, supports OTU clustering via hierarchical methods like the average neighbor algorithm.

3.2 Diversity and Statistical Analysis

R packages like phyloseq (McMurdie & Holmes, 2013) and vegan (Oksanen et al., 2012) provide robust platforms for calculating alpha and beta diversity using metrics such as Bray–Curtis, UniFrac, and Shannon indices. These are often visualized with ggplot2 (Wickham, 2016), enabling ordination plots (e.g., PCoA, NMDS), bar charts, and rarefaction curves.

3.3 Functional Annotation and Pathway Prediction

Bioinformatics has also enabled high-resolution functional annotation:

- PROKKA (Seemann, 2014), a Perl and C++ pipeline, integrates tools such as BLAST+ (Camacho et al., 2009), HMMER3 (Eddy, 2011), and Prodigal (Hyatt et al., 2010) to annotate microbial genomes.
- eggNOG-mapper v2 (Cantalapiedra et al., 2021), primarily in Python and R, uses DIAMOND (Buchfink et al., 2015) to assign genes to orthologous groups based on the eggNOG 5.0 database (Huerta-Cepas et al., 2017, 2019).
- The KEGG Mapper suite (Kanehisa & Goto, 2000; Kanehisa et al., 2015) contextualizes gene functions within metabolic and signaling pathways.

3.4 Taxonomic Classification

Taxonomic profiling is now achieved with tools offering both speed and precision:

- Kraken2 (Wood et al., 2019), written in C++, classifies sequences using exact k-mer matching and a lowest common ancestor (LCA) algorithm.
- MetaPhlAn3 (Beghini et al., 2021), based on clade-specific marker genes and Bowtie2 (Langmead & Salzberg, 2012), enables strain-resolved profiling.
- The RDP Classifier (Wang et al., 2007) uses a Naïve Bayesian algorithm for 16S-based classification.
- Reference databases such as SILVA (Quast et al., 2012), GTDB (Parks et al., 2018, 2021), and Greengenes (DeSantis et al., 2006) underpin most pipelines.

3.5 Assembly and Genome Binning

In shotgun metagenomics, reads are assembled and binned to recover metagenome-assembled genomes (MAGs):

- SPAdes (Bankevich et al., 2012) and metaSPAdes (Nurk et al., 2017), written in C++, use De Bruijn graph algorithms for short-read assembly.
- MEGAHIT (Li et al., 2015) employs succinct De Bruijn graphs for large-scale assemblies.
- Genome binning tools like MetaBAT2 (Kang et al., 2019) and MaxBin2 (Wu et al., 2016) group contigs using tetranucleotide frequency, coverage profiles, and Expectation-Maximization (EM) algorithms.

3.6 Visualization and Reproducibility

Visualizing results is critical to ecological interpretation:

- KronaTools (Ondov et al., 2011) generates interactive pie charts using HTML5/JavaScript for hierarchical taxonomy.
- R-based tools (phyloseq, vegan, ggplot2) allow for advanced statistical plotting.
- Reproducibility is enhanced through environments such as RStudio, Jupyter, and libraries like Biopython (Cock et al., 2009) and Bioconductor (Huber et al., 2015), which offer modules for sequence handling, annotation, and modeling.

Together, these high-throughput technologies and software tools form an integrated ecosystem that empowers researchers to investigate microbial diversity, structure, and function at unprecedented resolution across both natural and engineered environments.

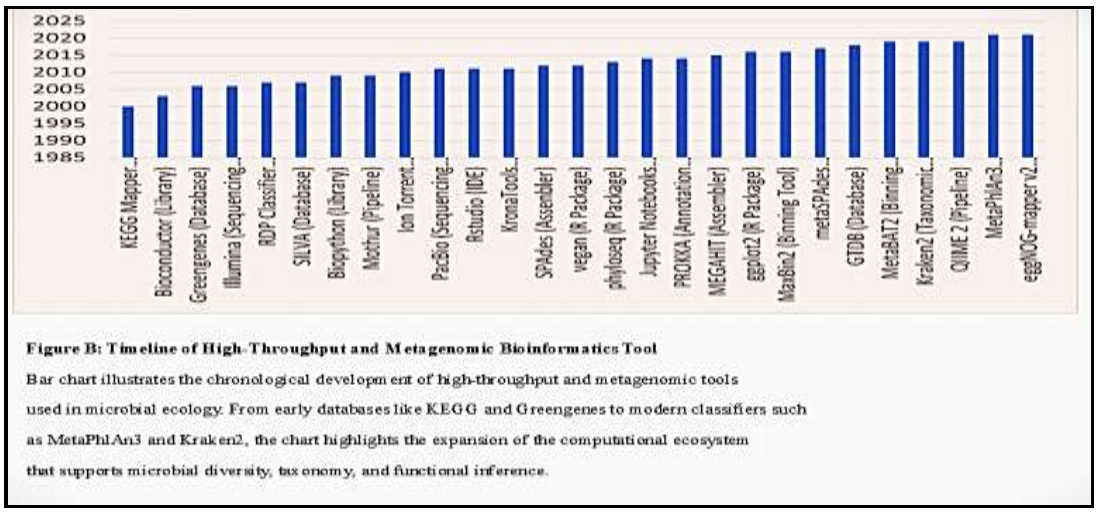

Figure B: Timeline of High-Throughput and Metagenomic Bioinformatics Tool

Bar chart illustrates the chronological development of high-throughput and metagenomic tools used in microbial ecology. From early databases like KEGG and Greengenes to modern classifiers such as MetaPhlAn3 and Kraken2, the chart highlights the expansion of the computational ecosystem that supports microbial diversity, taxonomy, and functional inference.

4. Application in Microbial Ecology and Diversity Studies

In microbial taxonomy and diversity studies, the term "strain" denotes genomic variants within a species, yet such intra-species variations typically involving single nucleotide polymorphisms (SNPs), insertions, deletions, or minor structural rearrangements generally affect less than 1–2% of conserved genes like the 16S rRNA and thus do not compromise taxonomic identification at genus or species levels (Janda & Abbott, 2007). This genomic stability underpins the robustness of bioinformatics-based identification, which relies on computational algorithms figure C such as the Naive Bayes Classifier (used in QIIME 2 (Bolyen et al., 2019) for taxonomic assignment), UCLUST (Edgar, 2010) and VSEARCH (Rognes et al., 2016) (for OTU clustering), and HMMER (for identifying conserved protein domains via Hidden Markov Models), all integrated within pipelines built in programming languages such as Python (e.g., Biopython, scikit-bio), R (e.g., phyloseq, vegan), C++ (e.g., Kraken2, MetaPhlAn, MEGAHIT), and Perl (e.g., BioPerl, Mothur). These workflows gain enhanced accuracy and consistency through integration with curated bioinformatics databases and portals, including the National Center for Biotechnology Information (NCBI) [55], which hosts curated genomic repositories such as GenBank and RefSeq (Sayers et al., 2022), the Kyoto Encyclopedia of Genes and Genomes (KEGG), known for its gene, enzyme, and pathway annotations (Kanehisa et al., 2021), the Clusters of Orthologous Groups (COGs) database, for functional protein classification (Galperin et al., 2015), Integrated Microbial Genomes & Microbiomes (IMG/M) for comparative microbial genomics (Chen et al., 2020), the SILVA rRNA database for aligned and quality-filtered ribosomal sequences (Quast et al., 2012), the Ribosomal Database Project (RDP), which offers taxonomic classifiers and 16S data (Cole et al., 2013), the Genome Taxonomy Database (GTDB), which provides a standardized, phylogeny-based taxonomy from genome data (Parks et al., 2018), RAST (Aziz et al., 2008), and MG-RAST (Metagenomics Rapid Annotations using Subsystems Technology) for functional annotation and visualization of metagenomes (Meyer et al., 2008). Together, these tools and platforms ensure that even in the face of strain-level mutations, the core taxonomic and functional features remain detectable and accurately classified, making bioinformatics an essential and resilient framework for microbial ecology, diversity, and evolutionary studies.

The application of bioinformatics in microbial ecology has evolved significantly beyond traditional metrics of species richness or abundance counts. Early approaches, such as direct enumeration or basic diversity indices like Shannon, Simpson, and Chao1, offered foundational insights into community composition. However, with the rise of high-throughput sequencing and computational ecology, more nuanced and scalable analytical methods have emerged, enabling multidimensional characterization of microbial communities.

Modern ecological analyses begin with the computation of alpha diversity using indices such as Shannon, Simpson, and Faith's Phylogenetic Diversity. These metrics quantify intra-sample microbial richness and evenness. These are routinely implemented through R packages like phyloseq (McMurdie & Holmes, 2013), vegan (Oksanen et al., 2012), and Python-based tools integrated in QIIME 2 (Bolyen et al., 2019). For inter-sample comparisons (beta diversity), dissimilarity metrics such as Bray-Curtis, Jaccard, and phylogeny-informed UniFrac distances are employed to assess ecological distances among microbial assemblages (Marotz et al., 2018). These measures form the foundation for ordination techniques, including Principal Component Analysis (PCA), Principal Coordinates Analysis (PCoA), and Non-Metric Multidimensional Scaling (NMDS) (Legendre & Birks 2012), which are used to reduce high-dimensional data and visually explore community structure. Further advancements have allowed researchers to infer ecological relationships and potential interactions within microbial consortia. Co-occurrence and co-exclusion network analyses, using tools such as CoNet (Faust et al., 2012) and SparCC (Friedman & Alm, 2012), enable the identification of community modules and keystone taxa microorganisms that disproportionately influence ecosystem structure or function. Visualization platforms like Cytoscape (Shannon et al., 2003) facilitate the interpretation of these complex networks, enhancing ecological inference.

Functional inference has also become a cornerstone of microbial ecological studies. Tools such as PICRUSt2 (Douglas et al., 2020), Tax4Fun (Aßhauer et al., 2015), and FAPROTAX (Louca et al., 2016) predict functional profiles from 16S rRNA marker gene data, allowing researchers to approximate metabolic potential without requiring full metagenomes. These predictions have unveiled trends in environmental adaptation, such as the presence of genes related to oxidative stress resistance, nutrient acquisition, and extremotolerance in various ecosystems.

In sum, bioinformatics has shifted microbial ecology from simple quantification to a system-level understanding of diversity, function, and interaction. These integrative approaches continue to reshape our capacity to interpret microbial communities in complex environmental settings.

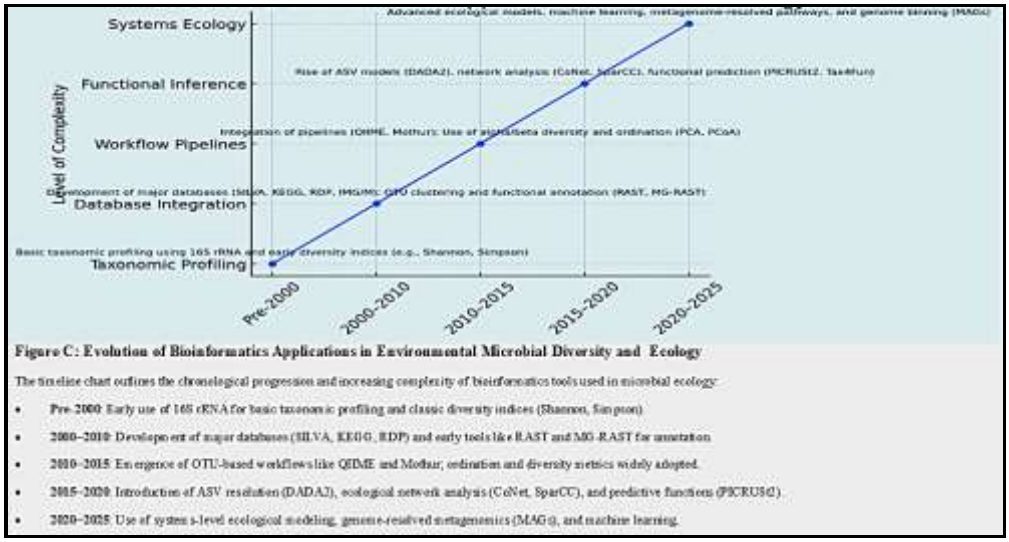

Figure C: Evolution of Bioinformatics Applications in Environmental Microbial Diversity and Ecology

The timeline chart outlines the chronological progression and increasing complexity of bioinformatics tools used in microbial ecology:

- **Pre-2000**: Early use of 16S rRNA for basic taxonomic profiling and classic diversity indices (Shannon, Simpson).
- **2000–2010**: Development of major databases (SILVA, KEGG, RDP) and early tools like BLAST and MG-RAST for annotation.
- **2010–2015**: Emergence of OTU-based workflows like QIIME and Mothur; ordination and diversity metrics widely adopted.
- **2015–2020**: Introduction of ASV resolution (DADA2), ecological network analysis (CoNet, SparCC), and predictive functions (PICRUSt2).
- **2020–2025**: Use of systems-level ecological modeling, genome-resolved metagenomics (MAGs), and machine learning.

5. Integrating Multi-Omics and Artificial Intelligence in Microbial Bioinformatics

5.1 Multi-Omics Integration

Modern microbial bioinformatics (Figure D) has progressed from single-marker analyses to comprehensive multi-omics strategies, enabling deeper insights into the functional, structural, and ecological dimensions of microbial communities.

Transcriptomics tools such as HISAT and HISAT2 (Kim et al., 2015; 2019) and DESeq2 (Love et al., 2014) allow for high-throughput alignment and differential expression analysis of RNA-Seq data, shedding light on active gene regulation under varying environmental conditions. In proteomics, platforms like MaxQuant (Cox & Mann, 2008) and PEAKS (Ma et al., 2003) enable accurate identification and quantification of proteins from mass spectrometry data, facilitating the study of cellular processes at the protein level.

In the field of metabolomics, tools such as MZmine (Pluskal et al., 2010) and MetaboAnalyst (Pang et al., 2021) support the detection, annotation, and interpretation of small-molecule metabolites, connecting microbial activity to phenotype and environmental adaptation.

These omics layers are increasingly integrated through systems biology platforms such as the DOE Systems Biology Knowledgebase (KBase) (Arkin et al., 2018), PATRIC (Wattam et al., 2016), and MG-RAST (Meyer et al., 2008), which support cross-platform analysis, comparative genomics, and functional annotation within unified frameworks. Visualization techniques such as heatmaps, correlation matrices, and Venn diagrams are routinely used to highlight co-regulated genes, metabolic overlaps, or taxonomic redundancies across datasets.

Applied Example: In hypersaline microbial mats, multi-omics integration has revealed gene expression profiles linked to Osmo-protection and sulfur metabolism, shedding light on microbial survival under extreme salinity and UV exposure.

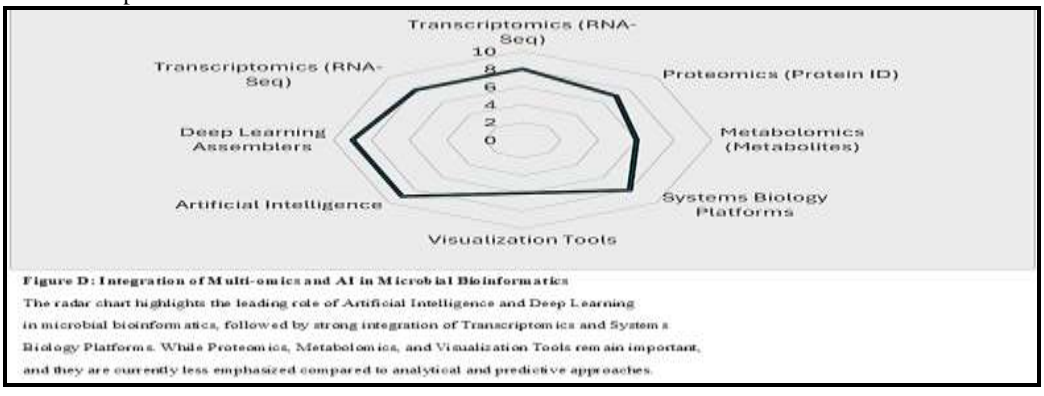

Figure D: Integration of Multi-omics and AI in Microbial Bioinformatics

The radar chart highlights the leading role of Artificial Intelligence and Deep Learning in microbial bioinformatics, followed by strong integration of Transcriptomics and Systems Biology Platforms. While Proteomics, Metabolomics, and Visualization Tools remain important, and they are currently less emphasized compared to analytical and predictive approaches.

5.2 Artificial Intelligence and Machine Learning in Microbiology

Simultaneously, artificial intelligence (AI) and machine learning (ML) are transforming the bioinformatics landscape by offering scalable, adaptive solutions to analyze high-dimensional microbial datasets.

These algorithms are increasingly employed to:

- Predict microbial phenotypes from genotypes (Marcos-Zambrano et al., 2021)
- Detect low-abundance taxa in complex metagenomes (Asgari et al., 2018)
- Cluster heterogeneous multi-omics profiles
- Optimize taxonomic classification pipelines for both known and novel species

For example, DeepMicrobes leverages deep learning for real-time taxonomic assignment from metagenomic reads (Liang et al., 2020), while advanced assemblers like SemiBin2 (Pan et al., 2023) apply self-supervised contrastive learning and neural networks to improve the recovery of metagenome-assembled genomes (MAGs) from fragmented datasets.

Applied Example: In antibiotic resistance surveillance, AI models trained on genomic and transcriptomic data have been used to accurately predict the presence and expression of resistance genes in clinical isolates enabling faster and more targeted treatment decisions (Ali et al., 2023).

These AI-enhanced platforms dramatically improve both speed and accuracy in microbial ecology and evolutionary studies, particularly when handling noisy or incomplete datasets. Collectively, the integration of multi-omics with AI-driven analytics represents a powerful, future-ready approach in microbial bioinformatics offering holistic perspectives on the genotype–phenotype–environment continuum.

6. Quantum Tools in Biology: A Frontier Perspective

With the emergence of quantum computing and quantum information theory, a new frontier is opening in life sciences. Quantum algorithms are poised to revolutionize microbial bioinformatics by significantly accelerating computationally intensive tasks such as protein folding prediction, quantum machine learning for pattern recognition in -omics datasets, and molecular interaction modeling with unprecedented accuracy (see Table 2).

Simultaneously, the field of quantum biology, a multidisciplinary area exploring quantum phenomena in biological systems, is gaining traction. Studies suggest that quantum coherence, entanglement, and tunneling may play roles in photosynthesis, enzyme catalysis, olfaction, and avian navigation, challenging traditional biochemical paradigms (Lambert et al., 2013). Although still in its early stages, integrating quantum technologies into microbiology and genomics holds the potential to overcome limitations of classical algorithms and unlock new layers of biological understanding (Bauer et al., 2020; Cao et al., 2019).

To support this paradigm shift, a diverse ecosystem of quantum programming languages and frameworks is being actively developed. While Python remains the dominant language due to its integration with scientific libraries and accessibility, several domain-specific quantum platforms have emerged each designed for distinct biological and computational applications (Figure E). These tools are not merely theoretical they are already being explored in molecular docking, genetic optimization, and even quantum-based diagnostics.

Several frameworks have emerged to support quantum computing applications in biology (Figure F). For instance, Qiskit enables quantum simulations for protein folding and drug discovery (Aleksandrowicz et al., 2019), while Cirq supports quantum neural networks tailored to biological data analysis (Google AI Quantum, 2020). PennyLane (Bergholm et al., 2018) and Strawberry Fields (Killoran et al., 2019), both developed by Xanadu, facilitating hybrid quantum-classical workflows and photonic quantum simulations, respectively. Additionally, QuTiP is Quantum Toolbox in Python (Johansson et al., 2012) commonly employed for modeling open quantum systems and quantum dynamics.

Although still theoretical, quantum computing holds potential for future applications in bioinformatics, particularly in the modeling of complex microbial systems, protein folding, and large-scale genomic data analysis. These possibilities remain under exploration but underscore the dynamic intersection of physics and biology in next-generation microbial research.

Table 2: Common Quantum Programming Languages and Frameworks Used in Biology

Tool / Language	Main Developer / Platform	Key Applications in Biology	Language Base	Reference
Qiskit	IBM Quantum	Quantum simulations, protein folding, drug discovery	Python	(Aleksandrowicz et al., 2019)
Cirq	Google Quantum AI	Quantum neural networks, quantum feature maps	Python	(Google AI Quantum, 2020)
PennyLane	Xanadu	Hybrid quantum-classical ML in -omics and genomics	Python	(Bergholm et al., 2018)
Q#	Microsoft Azure Quantum	Simulation of quantum circuits, quantum chemistry	Q#	(Svore et al., 2018)
Ocean SDK	D-Wave	Quantum annealing, biological network optimization	Python	D-Wave Systems Documentation (Official). https://docs.ocean.dwavesys.com
QuTiP	Open-source (community)	Quantum dynamics, biological coherence modeling	Python	(Johansson et al., 2012)
Strawberry Fields	Xanadu	Photonic quantum simulation of biological interactions	Python	(Killoran et al., 2019)

This table summarizes widely adopted quantum programming tools and platforms that are gaining relevance in biological and bioinformatics research. Each framework serves a specific purpose, from simulating protein folding and quantum chemistry (Qiskit, Q#) to optimizing biological networks (Ocean SDK) and modeling quantum coherence in cellular systems (QuTiP). Most of these tools are Python-based, allowing seamless integration with existing bioinformatics pipelines. Hybrid approaches, such as those enabled by PennyLane, bridge classical machine learning with quantum algorithms to enhance -omics data interpretation. As quantum biology and quantum machine learning continue to evolve, these languages represent foundational building blocks for future research in computational life sciences.

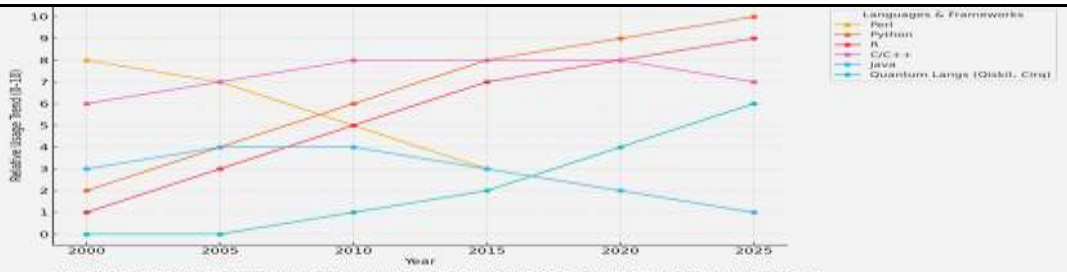

Figure E: Evolution of Programming Languages in Microbial Bioinformatics and Quantum Biology
Curves illustrating the evolution of key programming languages in microbial bioinformatics and their transition toward quantum biology applications, Perl dominated early bioinformatics (e.g., BioPerl) but has steadily declined. Python and R have seen a major rise due to their flexible libraries (e.g., Biopython, scikit-bio, phyloseq, vegan) and integration in modern pipelines (e.g., QIIME 2, machine learning). C/C++ remains consistently important for high-performance tools (e.g., Kraken2, MEGAHIT). Java shows decreasing use in this domain. Quantum Languages (e.g., Qiskit, Cirq, PennyLane) are emerging, reflecting the growing role of quantum machine learning and simulation in quantum biology and systems-level predictions.

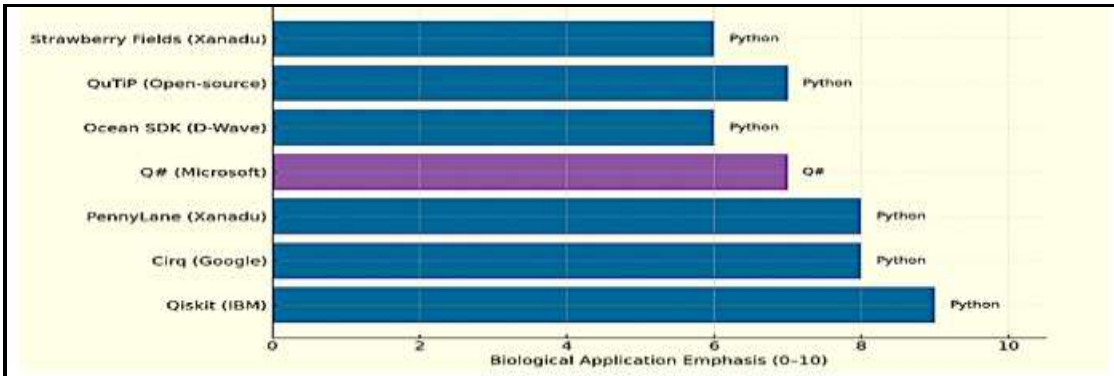

Figure F: Emerging Quantum Tools in Biology and Their Programming Language Base
Python-based frameworks dominate, reflecting Python's role as the central language for both classical and quantum bioinformatics. Qiskit, PennyLane, and Cirq lead in biological relevance, supporting tasks like quantum-enhanced -omics analysis, protein folding, and hybrid quantum-classical modeling. Q# (Microsoft) and Ocean SDK (D-Wave) are gaining ground, particularly in quantum chemistry and optimization of biological networks. QuTiP and Strawberry Fields show promise for modeling quantum coherence and photonic systems in biological settings.

7. CHALLENGES AND FUTURE DIRECTIONS

Despite the transformative impact of bioinformatics on microbial ecology and diversity research, several critical challenges persist. In many developing regions, limited access to computational infrastructure, high-performance servers, and licensed software continues to restrict the full deployment of advanced bioinformatics workflows (Bezuidenhout et al., 2017). Furthermore, the lack of standardized protocol spanning sequence processing,

taxonomic assignment, and functional prediction often hampers reproducibility and cross-study comparability (Ten Hoopen et al., 2017). (see Figure G)

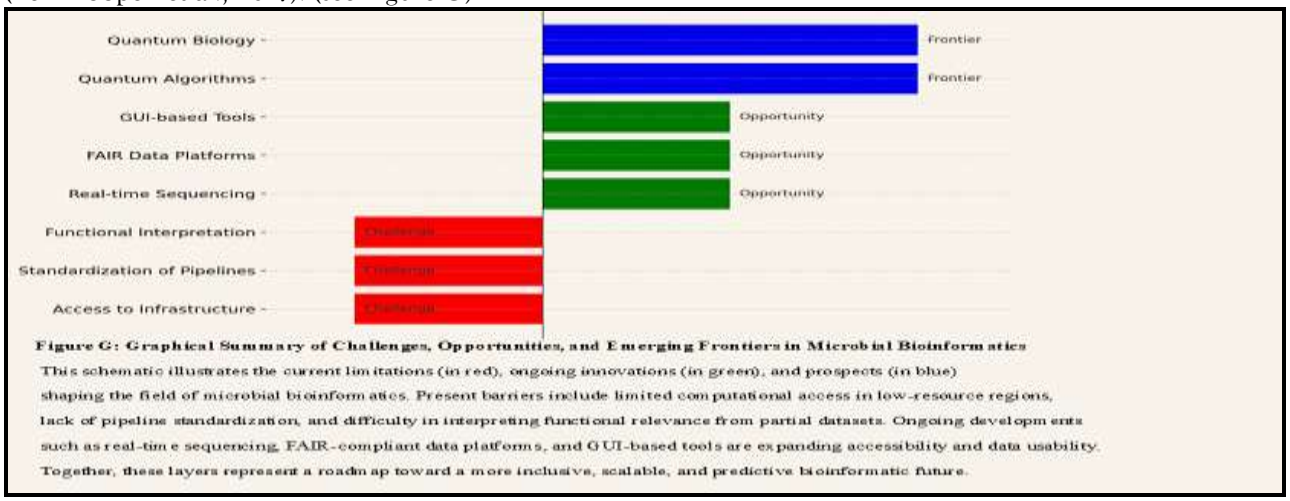

Figure G: Graphical Summary of Challenges, Opportunities, and Emerging Frontiers in Microbial Bioinformatics
This schematic illustrates the current limitations (in red), ongoing innovations (in green), and prospects (in blue) shaping the field of microbial bioinformatics. Present barriers include limited computational access in low-resource regions, lack of pipeline standardization, and difficulty in interpreting functional relevance from partial datasets. Ongoing developments such as real-time sequencing, FAIR-compliant data platforms, and GUI-based tools are expanding accessibility and data usability. Together, these layers represent a roadmap toward a more inclusive, scalable, and predictive bioinformatic future.

Another key limitation lies in the ecological interpretation of predicted genetic functions. Functional inferences derived from partial metagenomes or 16S rRNA gene data using tools like PICRUSt2 (Douglas et al., 2020) [69] and Tax4Fun (Aßhauer et al., 2015) depend heavily on reference databases, which may lack sufficient representation of taxa from underexplored or extreme ecosystems introducing bias or uncertainty.
Looking ahead, several emerging developments are poised to reshape the field:

- Real-time sequencing technologies like Oxford Nanopore are enabling in situ metagenomic monitoring, outbreak tracking, and environmental biosurveillance with increased speed and resolution (Loose et al., 2016).
- Community-curated, FAIR-compliant platforms such as the European Nucleotide Archive (ENA) (Mayer et al., 2021) and MGnify (Mitchell et al., 2019) are promoting open science, facilitating standardized metadata submission, and accelerating the reanalysis and reuse of microbiome datasets.
- The democratization of bioinformatics through user-friendly, GUI-based tools is increasingly empowering non-specialists in ecology, public health, and agriculture to perform sophisticated microbial analyses without requiring programming expertise (Mangul et al., 2019).

A particularly exciting and emerging frontier is the integration of quantum computing and quantum biology. Quantum algorithms are being explored to tackle computational bottlenecks in molecular simulations (Figure G), such as protein folding, genome-scale optimization, and pattern recognition across multi-omics datasets (Bauer et al., 2020; Cao et al., 2019). At a fundamental level, quantum biology seeks to explain phenomena such as enzyme catalysis, photosynthesis, and avian magnetoreception through mechanisms like quantum coherence and tunneling (Lambert et al., 2013). Although still in early stages, these concepts promise to shift our understanding of life toward a quantum-informed paradigm.
Ultimately, bridging existing gaps in accessibility, reproducibility, and ecological interpretation while embracing next-generation technologies such as quantum computing will be crucial to unlocking the full ecological and biotechnological potential of microbial communities in the decades to come.

8. CONCLUSION

Bioinformatics has transformed microbiology from a primarily observational field into a data-driven and predictive science. This shift is especially vital for studying microbial diversity in extreme and understudied environments where classical methods fall short. Advances ranging from sequence alignment to multi-omics integration and now artificial intelligence and quantum-informed models enable researchers to uncover not just microbial presence, but also function, adaptation, and ecological significance.
Integrating genomics with transcriptomics, proteomics, and metabolomics now provides systems-level insights, while machine learning redefines how we analyze complex ecological data. At the frontier, quantum computing offers promising solutions to model molecular and enzymatic processes with high precision, potentially overcoming bottlenecks in tasks like protein folding and genome search.

As bioinformatics tools become more accessible and standardized, the barriers to high-level microbiological research are diminishing. This opens opportunities for researchers worldwide, fostering broader participation in microbial science.

Looking ahead, bioinformatics will not only support but shape the future of microbial ecology and biotechnology. Whether it's discovering life in extreme environments or harnessing the quantum nature of biology, bioinformatics is more than a toolkit it is the language by which we explore and understand most of life on Earth.

REFERENCES

1. Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiological Reviews, 59(1), 143–169.
2. Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences, 74(11), 5088–5090.
3. Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal, 6(8), 1621–1624. https://doi.org/10.1038/ismej.2012.8
4. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., ... Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 7(5), 335–336. https://doi.org/10.1038/nmeth.f.303
5. Schloss, P. D., et al. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology, 75(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09
6. Keegan, K. P., Glass, E. M., & Meyer, F. (2016). MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function. Microbial Environmental Genomics (MEG), 207–233. https://doi.org/10.1007/978-1-4939-3369-3_13
7. Pal, S., Bhattacharya, M., Lee, S.-S., & Chakraborty, C. (2023). Quantum Computing in the Next-Generation Computational Biology Landscape: From Protein Folding to Molecular Dynamics. Molecular Biotechnology, 66(2), 163–178. https://doi.org/10.1007/s12033-023-00765-4
8. Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is Bioinformatics? A Proposed Definition and Overview of the Field. Methods of Information in Medicine, 40(04), 346–358. https://doi.org/10.1055/s-0038-1634431
9. Prescott, L. M., Harley, J. P., & Klein, D. A. (1996). Microbiology 3rd Edition, Wm. C.
10. Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673
11. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
12. Pearson, W. R., & Lipman, D. J. (1988). Improved Tools for Biological Sequence Comparison. Proceedings of the National Academy of Sciences, 85(8), 2444–2448. https://doi.org/10.1073/pnas.85.8.2444
13. Felsenstein, J. (1989). PHYLIP—Phylogeny Inference Package (Version 3.2). Cladistics, 5(2), 164–166.
14. Kumar, S., Tamura, K., & Nei, M. (1994). MEGA: Molecular Evolutionary Genetics Analysis Software for Microcomputers. Bioinformatics, 10(2), 189–191. https://doi.org/10.1093/bioinformatics/10.2.189
15. Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology, 73(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07
16. DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... & Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Applied and Environmental Microbiology, 72(7), 5069–5072. https://doi.org/10.1128/AEM.03006-05
17. Winsberg, E. (2010). Science in the Age of Computer Simulation. University of Chicago Press.
18. Illumina. (2023). *Sequencing systems overview*. https://www.illumina.com/systems/sequencing-platforms.html
19. Thermo Fisher Scientific. (n.d.). *Home*. https://www.thermofisher.com/dz/en/home.html
20. Pacific Biosciences. (n.d.). *SMRT sequencing technology*. https://www.pacb.com/smrt-science/smrt-sequencing/
21. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9
22. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461
23. Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ, 4, e2584. Portico. https://doi.org/10.7717/peerj.2584
24. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869
25. McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE, 8(4), e61217. https://doi.org/10.1371/journal.pone.0061217

26. Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Sólymos, P., Stevens, M. H. H., & Wagner, H. (2012). vegan: Community Ecology Package. Software http://CRAN.R-project.org/package=vegan
27. Wickham, H. (2016). Data Analysis. Ggplot2, 189–201. https://doi.org/10.1007/978-3-319-24277-4_9
28. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153
29. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., , K., & Madden, T. L. (2009). BLAST+: architecture and applications. BMC Bioinformatics, 10(1). https://doi.org/10.1186/1471-2105-10-421
30. Eddy, S. R. (2011). Accelerated Profile HMM Searches. PLoS Computational Biology, 7(10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195
31. Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics, 11(1). https://doi.org/10.1186/1471-2105-11-119
32. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution, 38(12), 58 25–5829. https://doi.org/10.1093/molbev/msab293
33. Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. Nature Methods, 12(1), 59–60. https://doi.org/10.1038/nmeth.3176
34. Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Molecular Biology and Evolution, 34(8), 2115–2122. https://doi.org/10.1093/molbev/msx148
35. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research, 47(D1), D309–D314. https://doi.org/10.1093/nar/gky1085
36. Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1), 27–30. https://doi.org/10.1093/nar/28.1.27
37. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research, 44(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070
38. Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biology, 20, 257. https://doi.org/10.1186/s13059-019-1891-0
39. Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. ELife, 10. CLOCKSS. https://doi.org/10.7554/elife.65088
40. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359. https://doi.org/10.1038/nmeth.1923
41. Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., & Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nature Methods, 12(10), 902–903. https://doi.org/10.1038/nmeth.3589
42. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., … & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research, 41(D1), D590–D596. https://doi.org/10.1093/nar/gks1219
43. Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nature Biotechnology, 36(10), 996–1004. https://doi.org/10.1038/nbt.4229
44. Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., & Hugenholtz, P. (2021). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research, 50(D1), D785–D794. https://doi.org/10.1093/nar/gkab776
45. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, 19(5), 455–477. https://doi.org/10.1089/cmb.2012.0021
46. Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Research, 27(5), 824–834. https://doi.org/10.1101/gr.213959.116
47. Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, 31(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033
48. Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ, 7, e7359. Portico. https://doi.org/10.7717/peerj.7359
49. Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2015). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics, 32(4), 605–607. https://doi.org/10.1093/bioinformatics/btv638
50. Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. BMC Bioinformatics, 12(1). https://doi.org/10.1186/1471-2105-12-385

51. Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163
52. Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nature Methods, 12(2), 115–121. https://doi.org/10.1038/nmeth.3252
53. Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. Journal of Clinical Microbiology, 45(9), 2761–2764. https://doi.org/10.1128/JCM.01228-07
54. National Center for Biotechnology Information (NCBI). (n.d.). *NCBI homepage*. U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov
55. Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2021). Database resources of the national center for biotechnology information. Nucleic Acids Research, 50(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112
56. Kanehisa, M., Sato, Y., Kawashima, M., Ishiguro-Watanabe, M., & Yamada, M. (2021). KEGG: integrating viruses and cellular organisms. Nucleic Acids Research, 49(D1), D545–D551. https://doi.org/10.1093/nar/gkaa970
57. Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Research, 43(D1), D261–D269. https://doi.org/10.1093/nar/gku1223
58. Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S., Varghese, N., Seshadri, R., Roux, S., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., & Kyrpides, N. C. (2020). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. Nucleic Acids Research, 49(D1), D751–D763. https://doi.org/10.1093/nar/gkaa939
59. Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R., & Tiedje, J. M. (2013). Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research, 42(D1), D633–D642. https://doi.org/10.1093/nar/gkt1244
60. Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., … Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics, 9(1). https://doi.org/10.1186/1471-2164-9-75
61. Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., & Edwards, R. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics, 9(1). https://doi.org/10.1186/1471-2105-9-386
62. Marotz, C. A., Sanders, J. G., Zuniga, C., Zaramela, L. S., Knight, R., & Zengler, K. (2018). Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome, 6(1). https://doi.org/10.1186/s40168-018-0426-3
63. Legendre, P., & Birks, H. J. B. (2012). From Classical to Canonical Ordination. Tracking Environmental Change Using Lake Sediments, 201–248. https://doi.org/10.1007/978-94-007-2745-8_8
64. Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., & Huttenhower, C. (2012). Microbial Co-occurrence Relationships in the Human Microbiome. PLoS Computational Biology, 8(7), e1002606. https://doi.org/10.1371/journal.pcbi.1002606
65. Friedman, J., & Alm, E. J. (2012). Inferring Correlation Networks from Genomic Survey Data. PLoS Computational Biology, 8(9), e1002687. https://doi.org/10.1371/journal.pcbi.1002687
66. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research, 13(11), 2498–2504. https://doi.org/10.1101/gr.1239303
67. Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. Nature Biotechnology, 38(6), 685–688. https://doi.org/10.1038/s41587-020-0548-6
68. Aßhauer, K. P., Wemheuer, B., Daniel, R., & Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics, 31(17), 2882–2884. https://doi.org/10.1093/bioinformatics/btv287
69. Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. Science, 353(6305), 1272–1277. https://doi.org/10.1126/science.aaf4507
70. Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature Methods, 12(4), 357–360. https://doi.org/10.1038/nmeth.3317
71. Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology, 37(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4
72. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biology, 15(12), 550. https://doi.org/10.1186/s13059-014-0550-8
73. Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology, 26(12), 1367–1372. https://doi.org/10.1038/nbt.1511
74. Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., & Lajoie, G. (2003). PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Communications in Mass Spectrometry, 17(20), 2337–2342. https://doi.org/10.1002/rcm.1196

75. Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics, 11, 395. https://doi.org/10.1186/1471-2105-11-395
76. Pang, Z., Zhou, G., Ewald, J., Chang, L., Hacariz, O., Basu, N., & Xia, J. (2022). Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. Nature Protocols, 17(8), 1735–1761. https://doi.org/10.1038/s41596-022-00710-w
77. Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., ... Yu, D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. Nature Biotechnology, 36(7), 566–569. https://doi.org/10.1038/nbt.4163
78. Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E. M., Disz, T., Gabbard, J. L., Gerdes, S., Henry, C. S., Kenyon, R. W., Machi, D., Mao, C., Nordberg, E. K., Olsen, G. J., Murphy-Olson, D. E., Olson, R., ... Stevens, R. L. (2016). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Research, 45(D1), D535–D542. https://doi.org/10.1093/nar/gkw1017
79. Marcos-Zambrano, L. J., Karaduzovic-Hadziabdic, K., Loncar Turukalo, T., Przymus, P., Trajkovik, V., Aasmets, O., Berland, M., Gruca, A., Hasic, J., Hron, K., Klammsteiner, T., Kolev, M., Lahti, L., Lopes, M. B., Moreno, V., Naskinova, I., Org, E., Paciência, I., Papoutsoglou, G., ... Truu, J. (2021). Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. Frontiers in Microbiology, 12. https://doi.org/10.3389/fmicb.2021.634511
80. Asgari, E., Garakani, K., McHardy, A. C., & Mofrad, M. R. K. (2018). MicroPheno: predicting environments and host phenotypes from 16S rRNA gene sequencing using a k-mer based representation of shallow sub-samples. Bioinformatics, 34(13), i32–i42. https://doi.org/10.1093/bioinformatics/bty296
81. Liang, Q., Bible, P. W., Liu, Y., Zou, B., & Wei, L. (2020). DeepMicrobes: taxonomic classification for metagenomics with deep learning. NAR Genomics and Bioinformatics, 2(1), lqaa009. https://doi.org/10.1093/nargab/lqaa009
82. Pan, S., Zhao, X.-M., & Coelho, L. P. (2023). SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing. Bioinformatics, 39(Supplement_1), i21–i29. https://doi.org/10.1093/bioinformatics/btad209
83. Ali, T., Ahmed, S., & Aslam, M. (2023). Artificial Intelligence for Antimicrobial Resistance Prediction: Challenges and Opportunities towards Practical Implementation. Antibiotics, 12(3), 523. https://doi.org/10.3390/antibiotics12030523
84. Lambert, N., Chen, Y.-N., Cheng, Y.-C., Li, C.-M., Chen, G.-Y., & Nori, F. (2012). Quantum biology. Nature Physics, 9(1), 10–18. https://doi.org/10.1038/nphys2474
85. Bauer, B., Bravyi, S., Motta, M., & Chan, G. K.-L. (2020). Quantum Algorithms for Quantum Chemistry and Quantum Materials Science. Chemical Reviews, 120(22), 12685–12717. https://doi.org/10.1021/acs.chemrev.9b00829
86. Cao, Y., Romero, J., Olson, J. P., Degroote, M., Johnson, P. D., Kieferová, M., Kivlichan, I. D., Menke, T., Peropadre, B., Sawaya, N. P. D., Sim, S., Veis, L., & Aspuru-Guzik, A. (2019). Quantum Chemistry in the Age of Quantum Computing. Chemical Reviews, 119(19), 10856–10915. https://doi.org/10.1021/acs.chemrev.8b00803
87. leksandrowicz, G., Alexander, T., Barkoutsos, P., Bello, L., Ben-Haim, Y., Bucher, D., ... & Zoufal, C. (2019). Qiskit: An open-source framework for quantum computing. https://doi.org/10.5281/zenodo.2562110
88. Google AI Quantum. (2020). Cirq: A Python framework for NISQ quantum circuits [Computer software]. GitHub. https://github.com/quantumlib/Cirq
89. Bergholm, V., Izsac, J., Schuld, M., Gogolin, C., Ahmed, S., Ajith, V., ... & Killoran, N. (2018). Pennylane: Automatic differentiation of hybrid quantum-classical computations. arXiv preprint arXiv:1811.04968
90. Killoran, N., Izaac, J., Quesada, N., Bergholm, V., Amy, M., & Weedbrook, C. (2019). Strawberry Fields: A Software Platform for Photonic Quantum Computing. Quantum, 3, 129. CLOCKSS. https://doi.org/10.22331/q-2019-03-11-129
91. Johansson, J. R., Nation, P. D., & Nori, F. (2012). QuTiP: An open-source Python framework for the dynamics of open quantum systems. Computer Physics Communications, 183(8), 1760–1772. https://doi.org/10.1016/j.cpc.2012.02.021
92. Bezuidenhout, L. M., Leonelli, S., Kelly, A. H., & Rappert, B. (2017). Beyond the digital divide: Towards a situated approach to open data. Science and Public Policy, 44(4), 464–475. https://doi.org/10.1093/scipol/scw036
93. ten Hoopen, P., Finn, R. D., Bongo, L. A., Corre, E., Fosso, B., Meyer, F., Mitchell, A., Pelletier, E., Pesole, G., Santamaria, M., Willassen, N. P., & Cochrane, G. (2017). The metagenomic data life-cycle: standards and best practices. GigaScience, 6(8). https://doi.org/10.1093/gigascience/gix047
94. Loose, M., Malla, S., & Stout, M. (2016). Real-time selective sequencing using nanopore technology. Nature Methods, 13(9), 751–754. https://doi.org/10.1038/nmeth.3930
95. Mayer, G., Müller, W., Schork, K., Uszkoreit, J., Weidemann, A., Wittig, U., Rey, M., Quast, C., Felden, J., Glöckner, F. O., Lange, M., Arend, D., Beier, S., Junker, A., Scholz, U., Schüler, D., Kestler, H. A., Wibberg, D., Pühler, A., ... Turewicz, M. (2021). Implementing FAIR data management within the German Network for Bioinformatics Infrastructure (de.NBI) exemplified by selected use cases. Briefings in Bioinformatics, 22(5). https://doi.org/10.1093/bib/bbab010
96. Mitchell, A. L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M. R., Kale, V., Potter, S. C., Richardson, L. J., Sakharova, E., Scheremetjew, M., Korobeynikov, A., Shlemov, A., Kunyavskaya, O., Lapidus, A., & Finn, R. D. (2019). MGnify: the microbiome analysis resource in 2020. Nucleic Acids Research. https://doi.org/10.1093/nar/gkz1035
97. Mangul, S., Mosqueiro, T., Abdill, R. J., Duong, D., Mitchell, K., Sarwal, V., Hill, B., Brito, J., Littman, R. J., Statz, B., Lam, A. K.-M., Dayama, G., Grieneisen, L., Martin, L. S., Flint, J., Eskin, E., & Blekhman, R. (2019). Challenges and recommendations to improve the installability and archival stability of omics computational tools. PLOS Biology, 17(6), e3000333. https://doi.org/10.1371/journal.pbio.3000333