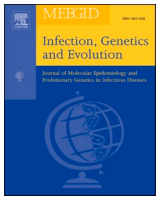
Infection, Genetics and Evolution 100 (2022) 105267

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



Review

Metagenomics and artificial intelligence in the context of human health

Atif Khurshid Wani [a], Priyanka Roy [b], Vijay Kumar [b,*], Tahir ul Gani Mir [a]

[a] Department of Biotechnology, School of Bioengineering and Biosciences, Lovely Professional University, Punjab 144411, India
[b] Department of Basic and Applied Sciences, National Institute of Food Technology Entrepreneurship and Management, Sonipat 131 028, Haryana, India

ARTICLE INFO

Keywords:
Microbiome
Metagenomics
Artificial intelligence
Human health
Diseases

ABSTRACT

Human microbiome is ubiquitous, dynamic, and site-specific consortia of microbial communities. The pathogenic nature of microorganisms within human tissues has led to an increase in microbial studies. Characterization of genera, like *Streptococcus*, *Cutibacterium*, *Staphylococcus*, *Bifidobacterium*, *Lactococcus* and *Lactobacillus* through culture-dependent and culture-independent techniques has been reported. However, due to the unique environment within human tissues, it is difficult to culture these microorganisms making their molecular studies strenuous. MGs offer a gateway to explore and characterize hidden microbial communities through a culture-independent mode by direct DNA isolation. By function and sequence-based MGs, Scientists can explore the mechanistic details of numerous microbes and their interaction with the niche. Since the data generated from MGs studies is highly complex and multi-dimensional, it requires accurate analytical tools to evaluate and interpret the data. Artificial intelligence (AI) provides the luxury to automatically learn the data dimensionality and ease its complexity that makes the disease diagnosis and disease response easy, accurate and timely. This review provides insight into the human microbiota and its exploration and expansion through MG studies. The review elucidates the significance of MGs in studying the changing microbiota during disease conditions besides highlighting the role of AI in computational analysis of MG data.

1. Introduction

The human microbiome is an amalgam of complete microbiota that is present in different anatomical sites (Ursell et al., 2012). Microorganisms existing on or within different body tissues may be in a commensal or mutualistic relationship (Khan et al., 2019). Several microorganisms reside in different sites of the body that harm the host by producing certain toxic metabolites like trimethylamine (Pickard et al., 2017). On the other hand, the human body also consists of several microorganisms that are opportunistic in nature (De Wit and Clumeck, 2017). They take advantage of the weak immune system and thus harm the immuno-compromised individuals. The conventional way of studying the microorganisms of a particular environment involves a culture-based strategy. But, it is often very difficult to provide the replica of natural microbial niche which makes exploration of previously unknown microbial members within the human microbiome difficult (Dekaboruah et al., 2020). MGs has expanded the horizon by the identification of microorganisms and microbial products like thermophilic *Bacillales* (Talamantes-Becerra et al., 2020; Wani et al., 2021) *and* antibiotic-resistant bacteria (Najar et al., 2020). Microbial profiles based on 16S *rRNA* genes have been used extensively for microbiome studies, but whole genome sequencing (WGS) is becoming more popular due to decreasing cost of sequencing (Schwarze et al., 2020). The function- and sequence-based MGs has led to the generation of new data sets that provide significant information about the medical aspect of the human microbiome (Handelsman, 2004). The human microbiome project (HMP), led by the National Institutes of Health (NIH), was one of the prime projects that evaluated the microbiota of the human body and it produced about 35 billion reads using 16S rRNA MG data from 690 samples across more than 10 sites (Peterson et al., 2009). The American Gut Project (McDonald et al., 2018) and Human Intestinal Tract (Qin et al., 2010) have significantly increased the human microbiome data with composition and function. These studies offer essential data to explore host-microbiome associations and their relation to disease development and progression. Most of the data generation occurs through amplicon sequencing using *16S rRNA* and *18S rRNA* marker gene allowing identification of new operational taxonomic units (OTUs) (Banos et al., 2018; Clarridge, 2004). Recent advances in sequencing technologies have increasingly replaced OTUs with amplicon sequence variants. The clustering of OTUs and denoising of ASVs is followed by

* Corresponding author.
E-mail address: drvijay.niftem@gmail.com (V. Kumar).

https://doi.org/10.1016/j.meegid.2022.105267
Received 19 August 2021; Received in revised form 3 March 2022; Accepted 4 March 2022
Available online 10 March 2022

classification and annotation (Callahan et al., 2017). Despite the effectiveness of these methodologies, it has certain flaws like limited information of microbial species present in lesser numbers, classification possible only up to genus level, and limited information of gene function (Poretsky et al., 2014). Short sequencing allows more precise classification up to strain level besides providing information about gene function, such as Gene Ontology (GO) construction and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Schmid et al., 2018). Advanced analytical tools are needed to explore and characterize biological datasets to study the peculiar nature of microbiome, composition, function, and heterogeneity. This is crucial for predicting the host-microbiome association helpful in disease diagnosis and building response strategies for the improvement of human health. The gut microbiome has been reported as an essential segment of personalized medicine. It is known to contribute to individual variability in health, disease, and therapeutics (Kashyap et al., 2017). The benefits offered by MGs like antibiotic discovery, forensic investigation, bioprospecting enzymes and early disease diagnosis depend on the efficient processing of a large amount of generated data (Bashir et al., 2014; Wani et al., 2022a, Wani et al., 2022b). AI is the ability of machines to perceive the environment and build response strategies for achieving goals. AI ensures rapid and accurate analysis of huge datasets and builds models through hierarchical concepts (Qu et al., 2019). AI ensures exploration of potential biological models *via* data generation, mathematic modeling and computational algorithms which is difficult by conventional methods (Egli et al., 2020). Of late, AI applications have been extended in data mining, medical diagnosis, robotics, and biometric recognition. Various studies have used AI tools to analyze the human microbiome, harvesting obscure knowledge to understand taxonomy, function (Cai and Sun, 2011), host phenotype inference in disease prediction (Bonder et al., 2012), and characterization of microbial signatures (Koohi-Moghadam et al., 2019).

This review gives a comprehensive outlook of AI and its application in analyzing the huge datasets generated by MG studies besides providing insights into the human microbiome in the context of different diseases.

2. MGs and the human microbiome

Considering a large number of microhabitats and a large number of interactions among microbial species that exist within the human body, the microbiota is often regarded as a dynamic ecological community (Pereira and Berry, 2017). Fig. 1 provides an overview of the human microbiome present in different body tissues. Microorganisms adapt very well to diverse environmental conditions through a series of cellular and molecular mechanisms, thus each microbial community within the human body has a unique structure in its localized environment (Culyba and Tyne, 2021;Wani et al., 2022a, Wani et al., 2022b). MGs reveals the hidden microbial diversity by directly collecting the DNA from the sites and subjecting it to either function-based or sequence-based screening (Lam et al., 2015). The former involves storing DNA into the surrogate host through cloning techniques, whereas the latter involves direct sequencing of extracted DNA. MGs provides a powerful way of exploring microbial diversity, thereby allowing detailed and unbiased investigation of members of the microbial communities, genes and gene products (Ahmad et al., 2019). This also ensures additional insight into the role of microbial communities in the context of different diseases.

2.1. Skin microbiota

Earlier dermatological microbiology was confined to culture-based approaches. But advancements in culture-independent approaches have led to the in-depth characterization of microbial communities (Grogan et al., 2019). In a study carried out on 20 skin sites for topographical and temporal diversity, more than 200 genera were identified with Actinobacteria (54%) as the dominating phyla (Grice et al., 2009). Even though skin microbiota such as *Staphylococcus spp.* has been explored by culture-based techniques but this technique underestimates

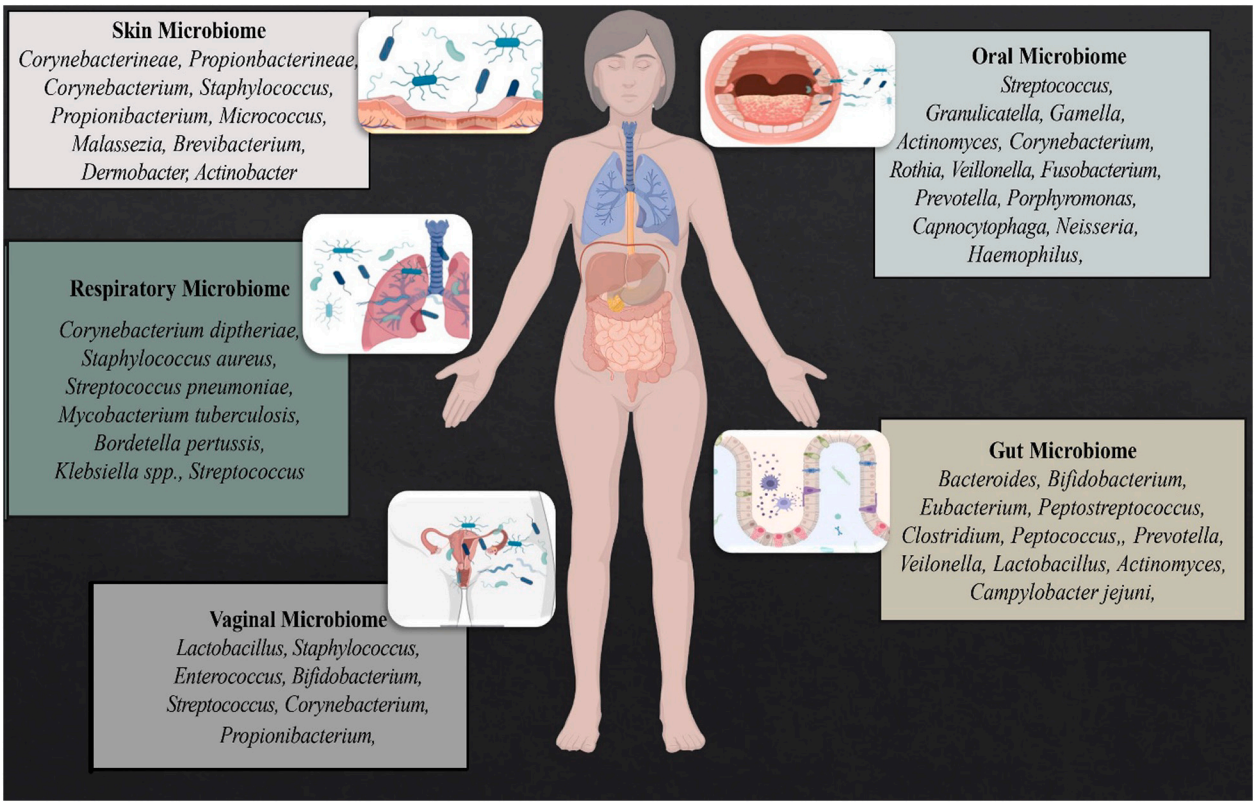


Fig. 1. Overview of the human microbiome localized in different body tissues.

the total diversity within the community (Boxberger et al., 2021). However, bacteria like *Corynebacterium spp.* and *Propionibacterium spp.* have been studied using culture-independent techniques to circumvent the culture bias (Kong, 2011). MGs approaches have prioritized the characterization of microorganisms associated with the diseases. *Propionibacterium acnes* phage and Propionibacterium abundance have been reported in healthy human skin through MGs but in acne patients, the *P. acnes* composition was found to be more diverse (Barnard et al., 2016). Several Fungal genera are also present on human skin, but in pathological conditions, some genera dominate over the other. For example, *Acremonium* and *Malassesia* are dominant during seborrheic dermatitis and atopic dermatitis, respectively (Faergemann, 2002). Table 1 provides an overview of the microbial diversity on or within different parts of the human body.

Table 1
Beneficial (B) and pathogenic (P) microorganisms found on or within different human body parts.

Microorganisms	Niche	Nature
Prevotella spp. (Tett et al., 2021)	Oral cavity, vagina, and gut	B (lipopolysaccharide and ammonia production that are part of vaginal mucus and epithelial cytokine production)
Veillonella parvula and V. dispar (Mashima et al., 2016)	Oral mucosa and intestines	B (confer strength and helps in digestion)
Streptococcus pyrogens (Helal et al., 2020; Slade and Vetter, 1956)	Maxillofacial region, upper respiratory tract, and skin	P (cellulitis, pharyngitis, and erysipelas)
Streptococcus dysgalactiae (Watanabe et al., 2017)	Skin, genital, and alimentary tract	P (endocarditis, meningitis, pneumonia)
Fusobacterium nucleatum and *F. nechrophoruum* (Kelly et al., 2018)	Oral cavity	P (periodontal and gingival infections)
Lactobacillus acidophilus and *Bifidobacterium bifidium* (Macfarlane and Cummings, 1999)	Gut	B (help in digestion)
Clostridium botulinum and *C. acetobutylium* (Camerini et al., 2019; Sreekumar et al., 2015)	Intestines	Both (production of botulinum toxin and generation of ethanol)
Lactobacillus rhamnosus (Schnabl, 2014)	Liver	B (helps in the prevention of alcoholic liver disease)
Escherichia coli (Abraham and Miao, 2015)	Kidney	Both (helps in absorption, but can block kidney's filters)
Moraxella catarrhalis (Okada et al., 2011)	Lungs	P (cause respiratory disorders)
Acinetobacter lwoffii (Qiu et al., 2011)	Lungs	B (reduces allergic airway inflammation)
Staphylococcus epidermidis (Khatoon et al., 2018)	Skin	P (formation of biofilms on plastic devices placed within the body)
Staphylococcus aureus (Tong et al., 2015)	Skin and upper respiratory tract	P (bacteremia and endocarditis)
Lactobacillus reuteri (Poutahidis et al., 2014)	Testicular tissue	B (increases spermatogenesis)
Corynebacterium mastitidis (Leger et al., 2017)	Eyes	B (stimulates cells to release antimicrobial factors)
Citrobacter freundii (Whalen et al., 2007; Puchenkova, 1996)	Sputum	Both (causes nosocomial infections and helps in reducing nitrate to nitrite)
Campylobacter upsaliensis (Facciolà et al., 2017)	Mouth	P (affects gastrointestinal tract)
Morganella morganii (Bangert et al., 1988)	Feces	B (filtration)
Mycoplasma orale (Fox et al., 1969)	Oropharynx	P in immunocompromised and non-pathogenic (NP) in immunocompetent individuals
Acinetobacter calcoacetius (Wong et al., 2017)	Whole-body distributed	Both

2.2. Gastrointestinal microbiota

Advancements in the MG sequencing techniques have opened ways for the exploration and characterization of estimated 10^4 microorganisms that dwell in the human gut (Wang et al., 2015). MG sequence-based study done on the samples of two healthy humans showed a tremendous enrichment of the microorganisms related to metabolic pathways of glycans, amino acids, xenobiotics, hormones and vitamins (Rowland et al., 2018). Researchers also applied the shotgun sequencing approach for the comparative microbiota study between lean and obese mice. They concluded that obese mice harbours microorganisms capable of encoding genes that cause the polysaccharide breakdown, which may not be possible to digest for a human or mice body (Thursby and Juge, 2017). It is important to mention that the microbiome composition in the GI tract changes with the diet (Conlon and Bird, 2014). Certain bacteria such as *Lactobacillus helveticus*, *Lactobacillus rhamnosus*, *Lactobacillus casei*, *L. plantarum*, *Bifidobacterium longum*, *B. infantis* and *B. brevis* have been reported to improve behavioural outcomes during autism, anxiety, depression, and obsessive-compulsive disorder due to their probiotic nature (Wang et al., 2016). However, the changes in the microbiota of GI are also related to harmful effects. A study reported that obese individuals have more Firmicutes and fewer Bacteroidetes in the gut (Magne et al., 2020).

2.3. Reproductive microbiota

Babies delivered by cesarean section foster less beneficial microbes and more pathogenic bacteria like *Staphylococcus aureus* and *Escherichia coli* (Neu and Rushing, 2011). However, babies born by vaginal delivery harbour more beneficial microbes (Neu and Rushing, 2011). Vagina-specific microorganisms protect human females from different infections (Larsen and Monif, 2001). *Lactobacillus* genus is found abundantly in premenopausal women. These help in fighting pathogenicity by producing lactic acid and hydrogen peroxide (Valenti et al., 2018). The microbial composition of the vagina also varies with the stages of the menstrual cycle (Chaban et al., 2014). Due to stigma vaginal microbiome studies has largely remained obscure through culture-based strategies. MGs cuts the barriers of conventional culture techniques and allow detailed studies of the microbiome from a unit sample which can be sequenced and amplified for the dynamic microbial studies of vaginal samples. A connection between preterm birth (PTB) and vaginal microbiota has been established in various research studies. One of the MG studies of the vaginal microbiome revealed significant differences in individual species among the women at higher risk of PTB and women at low risk of PTB (Feehily et al., 2020; Fettweis et al., 2019). *Lactobacillus crispatus* was found to be strongly linked with full-term pregnancies whereas, *Atopobium vaginae* and *Gardnerella vaginalis* were found to be associated with PTB (Fettweis et al., 2019). Off late, the uterus of a woman was believed to be a sterile region. However, studies have found the presence of various microorganisms inhabiting healthy women of reproductive age. However, the microbiome of the uterus shows a significant contrast from that of the GI tract and Vagina (Bardos et al., 2020; Franasiak and Scott, 2015). MG studies have revealed the presence of non-pathogenic, pathogenic and commensal bacteria, fungi and viruses in amniotic fluid and endometrium (Li et al., 2018). *Ureaplasma parvum, Streptomyces avermitllis, Neisseria lactamica* and *Candida spp.* are found as commensals within the uterus, whereas, *Mycoplasma hominis, Ureaplasma urealyticum* and *Candida spp.* have been found to be pathogenic (Plummer et al., 2021). Table 1 highlights some of the characterized microorganisms that are beneficial and/or harmful to human health.

3. MGs and production of large microbial datasets

A vast microbial diversity remains obscure because of the limitations in traditional microbiological culture techniques. MGs acts as a powerful

lens in exploring microbial diversity and ensure direct investigation of the useful genes to revolutionize the living world (Handelsman, 2004). The function- and sequence-based MGs has offered a great advantage in infection diagnosis by direct analysis of human-derived samples or from their environment (Markowitz et al., 2012). The adaptability of microorganisms in diverse niches and continuous fluctuations in the human microbiome during the diseases has made MGs and AI more relevant and necessary (Cao and Goodrich-Blair, 2017). The human microbiome comprises many microorganisms that consort with the host cell, thereby regulating physiological processes such as metabolism and immune response. Since the isolation of the first bacteria from feces, culture techniques have remained the primary methodology to characterize the members of the human microbiome (Browne et al., 2016). The microbial research got a significant push after the development of novel strategies such as the polymerase chain reaction (PCR) approach and 16S rRNA gene discovery. With MGs and Next-generation sequencing (NGS) technology, the direct sample analysis to examine functional genes for phylogenetic analysis become possible (Oulas et al., 2015). Several computational tools have also been developed for in-depth analysis of large data sets for accurate characterization and broader description of the microbial world. MetaHIT (METAgenomics of the Human Intestinal Tract) has led to the mining of MGs datasets for simplifying the variability and complexity associated with it (Qin et al., 2010). Fig. 2 illustrates the metagenomic strategies that result in the production of huge microbial datasets. The exponentially growing MG sequence data is highly complex; thus, its analysis is often a daunting challenge (Cheng et al., 2019). Metadata includes information about the phylogeny and cellular features of a particular microorganism living within the host and its environmental features (Cheng et al., 2019; Franco-Duarte et al., 2019). There have been several tools to analyze, integrate and characterize sequence data and metadata. In 2007, a team at the University of Chicago and Argonne National Laboratory released MG-RAST (Metagenomics Rapid Annotation Subsystem Technology) for the metagenomic data analysis (Meyer et al., 2008). Till 2012,14x10^{12}bases (14.8 terabases) of DNA were analyzed with almost 10,000 metagenome data sets available freely for comparative analysis with MG-RAST (Keegan et al., 2016). Moreover, IMG/M (Integrated Microbial Genomes/MGs) system also offers a pathway for metagenomic function analysis based on the sequence of the microbial communities (Markowitz et al., 2012). MEGAN (MEta Genome Analyzer) is a standalone software that is also used for high-throughput MG analysis since 2005 performing both functional as well as taxonomic binning (Huson, 2015; Huson et al., 2018).

4. AI learning methods for microbiome analysis

Various domains of AI are centred around specific goals including reasoning, representation, learning, processing, and perception. AI researchers have integrated and adapted numerous problem-solving tools such as mathematical optimization, artificial neural networks (ANNs), statistical methods, and logistic regression (LR). The tools evaluate models based on the input data. Some of the principal strategic tools are briefly discussed below:

4.1. *Artificial neural networks (ANNs)*

ANNs are perhaps the most sensitive and efficient strategies of ML. It bears a strong resemblance to the neural network of the human brain. A "neuron" in AI is a mathematical function network that collects and classifies data according to a particular architecture. The network

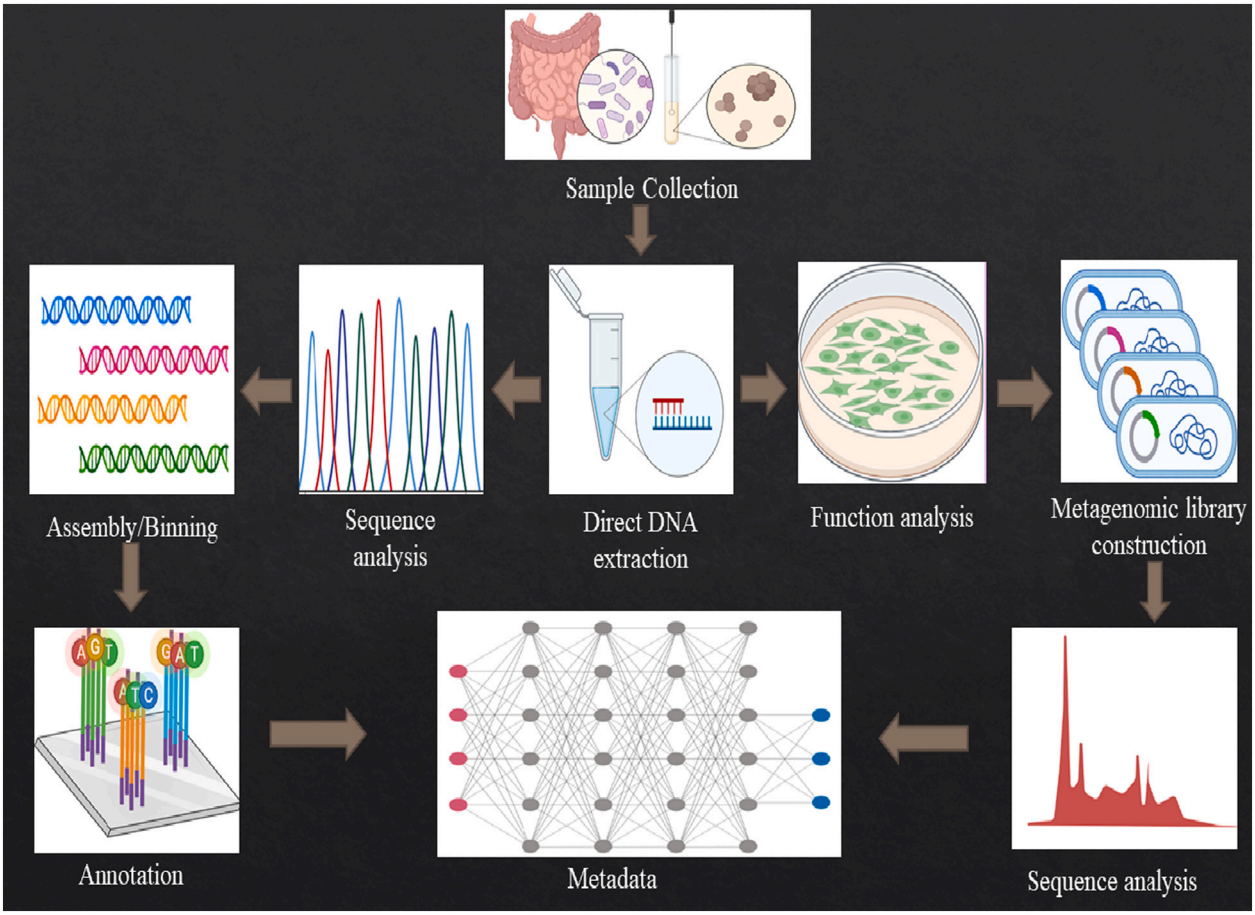


Fig. 2. Function- and sequence-based metagenomic strategies to produce large data sets essential for revealing microbial diversity and their potential nature.

functions like regression analysis of statistical methods. Bzhalava *et al*, trained RF and ANN using MG sequences taxonomically categorized into virus and non-virus classes. The accuracy reported was well beyond the chance level with the receiver operating characteristic curve of 0.79 (Bzhalava et al., 2018). In another study, scientists described a neural network tool for the characterization of host phenotypes from MG data, using a novel augmentation method for mitigating data over-fitting (Lo and Marculescu, 2019). They tested the proposed protocol on the datasets of hierarchical matching pursuit (Turnbaugh et al., 2009) and diseases like esophagitis, adenocarcinoma, and IBD (Marcos-Zambrano et al., 2021).

4.2. Logistic regression (LR)

LR is a statistical tool that utilizes logistic functions to model a dependent variable. It is used to understand the correlation between dependent and independent variables by probability estimation through the LR equation. Microbial signatures have been established in bacterial vaginosis using LR (Beck and Foster, 2015). Tap *et al*, used ML protocol to characterize microbial signature for predicting IB syndrome using the LASSO-LR approach (Tap et al., 2017). A similar approach was used by another research group for the extraction of a featured bacterial group as a biomarker in IB patients (Fukui et al., 2020). Wu *et al*, showed that the gut microbiome could be used to differentiate abnormal cases from controls with higher specificity. They concluded that microbiome biomarkers offer a promising predictive potential for disease diagnostics (Wu et al., 2018).

4.3. Hierarchical clustering (HC)

HC, an unsupervised ML technique, develops a hierarchy of clusters by splitting or merging clusters based on different metrics (Zhou and Gallins, 2019). Cai and Sun used HC for the classification of 16S rDNA and demonstrated its usefulness by addressing computational issues (Cai and Sun, 2011). HC has also been applied for establishing relations between diseases and microbiota. Richards *et al*, used transcriptome signatures of colonocytes and 16S rRNA data suggesting alterations in the microbiome can be useful in disease diagnostics and therapies (Richards et al., 2019).

5. Artificial intelligence: modus operandi for detecting `pathogens

Soon after the inception of AI in 1956, it started to gain attention from different research sectors and perhaps it was among the most valuable inventions of the 21st century. By its fast computational ability and simplified algorithms, enormous metadata can be generated quickly (Schmidt et al., 2019). To evaluate these large data sets of the microbial world AI ensures deep learning and understanding in terms of hierarchical concepts (Agrebi and Larbi, 2020). Machine learning (ML) is a part of AI that involves the study of computer algorithms to build a model based on input data that are used in computational statistics making predictions easy and accurate. ML in synchrony with AI explores models, potential patterns, adaptability, behaviours, niche associations to determine the microbial lifestyle within a particular environment (Palma et al., 2018). Table 2 highlights some of the ML tools used in microbiome analysis for human welfare. This has led to novel strain recognition, process understanding and medical diagnosis (Ghannam and Techtmann, 2021). Since the discovery of pathogens involves stepwise arduous and time-taking research processes, amid the COVID-19 pandemic the rapid discovery through computational tools has become essential. As MG sequencing, especially whole-genome sequencing (WGS), is turning out to be valuable in clinical microbiology and other microbial research the need for data curation and interpretation is also indispensable. The attributes associated with WGS are complex nature, high throughput, large data volume, and multiple redundancies (Ji et al., 2017). AI offers a solution to these intricate problems by simplifying complex algorithms. The development of supercomputers, biosensors and remote servers allow the luxury of intensive microbial analysis with a relatively lower cost (National Research Council US Committee, 2000). In 2012, a study used AI tools to differentiate microbial diversity in Inflammatory bowel disease (IBD) and non-IBD persons. They used random-forest (RF) and supervised classification models to compare diverse algorithms, bagging and stacking and concluded that it could be used as an effective diagnostic tool with adequate predictive power (Li and Qian, 2020). RF is an ensemble learning strategy for classification and regression to develop uniformity in the data sets through the development of decision trees. Supervised learning (SL) is the ML task of studying function maps based on the input-output pairs. SL algorithms analyze data sets and develop inferred functions used for mapping and generalizing new algorithms. To carry out the comparative analysis of the MG profiling among diverse microbial communities harbouring in different environments, there are two approaches of AI that have been proposed. One is an assembly-based method of reduced representation while the other is read based taxonomy profiling (Pasolli et al., 2016; Zhong et al., 2021). Among several AI approaches RF technique showed accurate and promising results as an efficient classifier. The RF models developed from taxonomic profiling (read based) could reach up to the accuracy of more than 90% with 95% confidence (Boutaba et al., 2018). This signifies assembly-based and read-based strategies are substantial tools for metadata analysis. Using AI, MG samples can be traced accurately, and the composition of microbes can be simplified by AI algorithms.

Table 2
Some AI tools for microbiome analysis useful for taxonomic profiling, pathogen detection and disease prediction.

AI/ML tool	Applications
Meta-Signer (Reiman et al., 2021)	Feature ranking through Ensemble learning and metagenome signature identifier
DeepMicro (Oh and Zhang, 2020)	Deep representation learning for infection/ disease prediction using the microbiome data
mAML (Yang and Zou, 2020)	Automated human disease classification through reproducible and reproducible models
PaPrBaG (Deneke et al., 2017)	Reliable at low genomic coverage for pathogenicity prediction
MicrobiomeAnalystR (Dhariwal et al., 2017)	Comprehensive functional, statistical, and meta-analysis.
mothur (Chappidi et al., 2019)	Handling of multiple microbial datasets for community analysis
QIIME (Estaki et al., 2020)	End-to-end analysis of microbiome data
BiomMiner (Shamsaddini et al., 2020)	Exploratory microbiome analysis through auto-tuning of optimal parameters for visualization of clinical datasets
Scikit-learn (Feurer et al., 2019)	Robust ML system for predictive analysis
MIPMLP (Jasner et al., 2021)	Microbiome preprocessing through ML for accurate data analysis

5.1. Tracing outbreaks by AI

Outbreak tracing requires epidemiological details to reveal the infection source (Straif-Bourgeois et al., 2014). The WGS of the infection-causing microbes is extended to more microbial communities, which offer the possibility to leverage the microbial metadata. *Salmonella enterica*, a causative agent of food poisoning, has a wide range of persistence in different environments, though sublineages are mostly present within specific hosts (Kan et al., 2018). Recently, AI has been used for the molecular marker identification in *Salmonella typhimurium*, thereby ensuring the tracing of infections (Sudhakar et al., 2021). Zhang *et al*, worked on the problem of predicting the source using diverse data set and ML algorithms. They collected the *S. Typhimurium* genome from different sources including isolates from several zoonotic outbreaks. They utilized the RF model on the core genome variation and accessory genome of animal isolates only. 83% algorithm accuracy was achieved

and importantly, seven outbreaks were attributed to the correct source among eight isolates (Zhang et al., 2019). Lupolova *et al*, used a vector machine for predicting the host of *S. typhimurium* based on accessory genome. This model made it possible to differentiate avian, swine, bovine and human strains with 81% of algorithmic accuracy (Lupolova et al., 2017). These studies exemplify the importance of AI and ML algorithms in pathogen surveillance, thereby opening the gateway for predicting outbreaks.

5.2. Clinical utility of AI for detection of pathogens

The usefulness of AI in clinical microbiology has solved many problems and its applicability is imperative in the age of supercomputers and ML. The attributes like testing time, quality outcome and cost-effective approaches associated with AI have ensured rapid turn-around from conventional ways to AI ways (Siddesh et al., 2021). Multiple software tools have been developed through machine learning algorithms for clinical use like, 'Platform' – used for the classification of an anti-nuclear antibody (ANA) (Bizzaro et al., 2014) and 'CellaVision' – used for the malarial parasite detection in leukocytes by identifying the morphological differences (Florin et al., 2018). In one of the recent studies, the validated AI tools have been used for the detection of faecal protozoa by trichrome-stained faecal smears. AI has also been used in the rapid detection of methicillin-resistant *S. aureus* (Faron et al., 2016b), vancomycin-resistant *Enterococcus* (Faron et al., 2016a) and acid-fast bacilli (Rhoads, 2020).

5.3. Application of AI for the detection of food-agro and water-borne pathogens

Microbial diversity plays its role in agricultural yield by inducing positive and negative effects (Stefan et al., 2021). The positive effects include an increase in stress tolerance (Meena et al., 2017), nutrient uptake, and growth promotion (Pattnaik et al., 2021) while later is mostly associated with microbial pathogenicity (Brader et al., 2017). They may either enhance the yield or cause its reduction by causing several plant diseases. The main cause of contamination in agriculture-derived foods is the presence of pathogenic microorganisms in water (Cabral, 2010). The presence of enteric pathogens in agricultural water induces illness and thereby survival becomes difficult. *Salmonella spp.* are the vastly studied water-borne pathogens that affect the plant and ultimately the humans (Liu et al., 2018). Shiga-toxin producing *E. coli* has also been reported in the agricultural water of Florida (Topalcengiz and Danyluk, 2019). AI approaches have been successfully used in predicting the microbial load in the water. The indicator microorganisms, microbial load, faecal pollution can be checked by using artificial neural networks (ANNs). The combination of MGs, AI and ANNs helps in predicting the concentration of total coliform, faecal coliform, enterococci, steam flow, conductivity, precipitation, and turbidity of the agricultural surface water (Zhou et al., 2019). AI has also been used in the detection of protozoa such as cryptosporidium and giardia in water thus making it an important tool for safe and efficient agricultural production (Abeywardena et al., 2015).

6. AI and MGs: the way forward

MGs has opened a gateway towards studying microbial systems directly from the environment to gain a deeper understanding of the complex processes. This can be done by understanding evolutionary dynamics, phylogeny, and evaluation of toxicity. Since MGs allows scientists to access the metabolic and functional diversity but is unable to show the exact mechanism involved in active processes (Pearman et al., 2020). The metatranscriptome analysis is thus helpful in providing information about the complex regulation and expression profiles of these microbial communities (Bashiardes et al., 2016). However, the technical glitches and data evaluation is often very challenging for accurate genomic information of a particular microorganism and thus delay in diagnosis of human pathogens is always a drawback in the treatment of a disease. The AI-based modeling system offers an advantage in the fast, accurate and efficient evaluation and characterization of the microbial datasets. The cellular automata, ANNs and micro-evolution techniques can prove to be blessings in simulating the microbial population dynamics using simulators like biogenesis and neurokernel driven by executable DNA. Currently, multiplex rapid PCR are used in clinical diagnosis in most of the hospitals due to their easy handling, simple data interpretation, and lower cost with analytical time of about 2 h. This makes the use of shotgun MG approach unlikely in clinical utility. Since PCR models can only detect the commonly encountered pathogens, therefore shotgun MGs combined with AI tools can be effective in early infection diagnosis and identification of emerging pathogenic threats. In this review, we have attempted to highlight the combinatorial use of MGs and AI for disease diagnosis in the context of human microbiome and other pathogenic threats. With the increase in concern regarding disease outbreaks, MGs and AI will only become more demanding for microbial characterization and data analysis respectively.

7. Concluding remarks

AI is already showing a profound impact on the healthcare system, and it is expected that its applicability will gain further momentum. MGs followed by AI data evaluation can change the way we handle and analyze the healthcare data-sets including the ones that are related to microorganisms and the associated pathogenicity. Among the major advantages, the prediction of outbreaks through MGs and AI systems can prove to be a blessing to keep humanity safe. The understanding of hidden functional microbial diversity within the human body and surroundings can be rapidly achieved through the combination of MGs and AI. However, the challenge that exists is that the microbiologists and biotechnologists need to get familiar with data handling, digital microbiology and interpret the AI results and image analysis. The COVID period has further stamped the usefulness of digital microbiology in diverse health care systems from disease diagnosis to vaccination process the workflow can be made easy using an AI system.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Funding

Not applicable.

CRediT authorship contribution statement

Atif Khurshid Wani: Conceptualization, Literature review, Writing original draft, Data curation,Writing review & editing. **Priyanka Roy:** Investigation, Formal analysis, Methodology, Writing review & editing. **Vijay Kumar:** Project adminstration,Supervision, Validation, Funding acquisition, Resources. **Tahir ul Gani Mir:** Software, Formal analysis, Visualization, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

The authors/co-authors are thankful to the National Institute of Food Technology Entrepreneurship and Management (NIFTEM), Kundli, Sonipat, Haryana-131028, and School of Bioengineering and Biosciences, Lovely Professional University, Punjab (144411), India for supporting this project by providing necessary facilities.

References

Abeywardena, H., Jex, A.R., Gasser, R.B., 2015. A perspective on Cryptosporidium and Giardia, with an emphasis on bovines and recent epidemiological findings. Adv. Parasitol. 88, 243–301. https://doi.org/10.1016/bs.apar.2015.02.001.

Abraham, S.N., Miao, Y., 2015. The nature of immune responses to urinary tract infections. Nat. Rev. Immunol. 15, 655–663. https://doi.org/10.1038/nri3887.

Agrebi, S., Larbi, A., 2020. Use of artificial intelligence in infectious diseases. Artif. Intell. Precis. Health 415–438. https://doi.org/10.1016/B978-0-12-817133-2.00018-5.

Ahmad, T., Singh, R.S., Gupta, G., Sharma, A., Kaur, B., 2019. Chapter 15 - Metagenomics in the search for industrial enzymes. In: Singh, R.S., Singhania, R.R., Pandey, A., Larroche, C.B.T.-A. in E.T (Eds.), Biomass, Biofuels, Biochemicals. Elsevier, pp. 419–451. https://doi.org/10.1016/B978-0-444-64114-4.00015-7.

Bangert, R.L., Ward, A.C., Stauber, E.H., Cho, B.R., Widders, P.R., 1988. A survey of the aerobic bacteria in the feces of captive raptors. Avian Dis. 32, 53–62.

Banos, S., Lentendu, G., Kopf, A., Wubet, T., Glöckner, F.O., Reich, M., 2018. A comprehensive fungi-specific 18S rRNA gene sequence primer toolkit suited for diverse research issues and sequencing platforms. BMC Microbiol. 18, 190. https://doi.org/10.1186/s12866-018-1331-4.

Bardos, J., Fiorentino, D., Longman, R.E., Paidas, M., 2020. Immunological role of the maternal uterine microbiome in pregnancy: pregnancies pathologies and alterated microbiota. Front. Immunol. 10 https://doi.org/10.3389/fimmu.2019.02823.

Barnard, E., Shi, B., Kang, D., Craft, N., Li, H., 2016. The balance of metagenomic elements shapes the skin microbiome in acne and health. Sci. Rep. 6, 39491. https://doi.org/10.1038/srep39491.

Bashiardes, S., Zilberman-Schapira, G., Elinav, E., 2016. Use of metatranscriptomics in microbiome research. Bioinforma. Biol. Insights 10, 19–25. https://doi.org/10.4137/BBI.S34610.

Bashir, Y., Pradeep Singh, S., Kumar Konwar, B., 2014. Metagenomics: an application based perspective. Chin. J. Biol. 2014, 146030. https://doi.org/10.1155/2014/146030.

Beck, D., Foster, J.A., 2015. Machine learning classifiers provide insight into the relationship between microbial communities and bacterial vaginosis. BioData Min. 8, 23. https://doi.org/10.1186/s13040-015-0055-3.

Bizzaro, N., Antico, A., Platzgummer, S., Tonutti, E., Bassetti, D., Pesente, F., Tozzoli, R., Tampoia, M., Villalta, D., Study Group on Autoimmune Diseases of the Italian Society of Laboratory Medicine, Italy, 2014. Automated antinuclear immunofluorescence antibody screening: a comparative study of six computer-aided diagnostic systems. Autoimmun. Rev. 13, 292–298. https://doi.org/10.1016/j.autrev.2013.10.015.

Bonder, M.J., Abeln, S., Zaura, E., Brandt, B.W., 2012. Comparing clustering and preprocessing in taxonomy analysis. Bioinformatics 28, 2891–2897. https://doi.org/10.1093/bioinformatics/bts552.

Boutaba, R., Salahuddin, M.A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., Caicedo, O.M., 2018. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. J. Internet Serv. Appl. 9, 16. https://doi.org/10.1186/s13174-018-0087-2.

Boxberger, M., Cenizo, V., Cassir, N., La Scola, B., 2021. Challenges in exploring and manipulating the human skin microbiome. Microbiome 9, 125. https://doi.org/10.1186/s40168-021-01062-5.

Brader, G., Compant, S., Vescio, K., Mitter, B., Trognitz, F., Ma, L.-J., Sessitsch, A., 2017. Ecology and genomic insights into plant-pathogenic and plant-nonpathogenic endophytes. Annu. Rev. Phytopathol. 55, 61–83. https://doi.org/10.1146/annurev-phyto-080516-035641.

Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., Goulding, D., Lawley, T.D., 2016. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature 533, 543–546. https://doi.org/10.1038/nature17645.

Bzhalava, Z., Tampuu, A., Bała, P., Vicente, R., Dillner, J., 2018. Machine learning for detection of viral sequences in human metagenomic datasets. BMC Bioinforma. 19, 336. https://doi.org/10.1186/s12859-018-2340-x.

Cabral, J.P.S., 2010. Water microbiology. Bacterial pathogens and water. Int. J. Environ. Res. Public Health 7, 3657–3703. https://doi.org/10.3390/ijerph7103657.

Cai, Y., Sun, Y., 2011. ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time. Nucleic Acids Res. 39 https://doi.org/10.1093/nar/gkr349 e95.

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11, 2639–2643. https://doi.org/10.1038/ismej.2017.119.

Camerini, S., Marcocci, L., Picarazzi, L., Iorio, E., Ruspantini, I., Pietrangeli, P., Crescenzi, M., Franciosa, G., 2019. Type E botulinum neurotoxin-producing *Clostridium butyricum* strains are aerotolerant during vegetative growth. mSystems 4. https://doi.org/10.1128/mSystems.00299-18.

Cao, M., Goodrich-Blair, H., 2017. Ready or not: microbial adaptive responses in dynamic symbiosis environments. J. Bacteriol. 199 https://doi.org/10.1128/JB.00883-16.

Chaban, B., Links, M.G., Jayaprakash, T.P., Wagner, E.C., Bourque, D.K., Lohn, Z., Albert, A.Y., van Schalkwyk, J., Reid, G., Hemmingsen, S.M., Hill, J.E., Money, D.M., 2014. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. Microbiome 2, 23. https://doi.org/10.1186/2049-2618-2-23.

Chappidi, S., Villa, E.C., Cantarel, B.L., 2019. Using mothur to determine bacterial community composition and structure in 16S ribosomal RNA datasets. Curr. Protoc. Bioinformatics 67. https://doi.org/10.1002/cpbi.83 e83.

Cheng, M., Cao, L., Ning, K., 2019. Microbiome big-data mining and applications using single-cell technologies and metagenomics approaches toward precision medicine. Front. Genet. 10 https://doi.org/10.3389/fgene.2019.00972.

Clarridge, J.E., 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin. Microbiol. Rev. 17, 840–862. https://doi.org/10.1128/CMR.17.4.840-862.2004.

Conlon, M.A., Bird, A.R., 2014. The impact of diet and lifestyle on gut microbiota and human health. Nutrients 7, 17–44. https://doi.org/10.3390/nu7010017.

Culyba, M.J., Tyne, D.V., 2021. Bacterial evolution during human infection: adapt and live or adapt and die. PLoS Pathog. 17 https://doi.org/10.1371/journal.ppat.1009872 e1009872.

De Wit, S., Clumeck, N., 2017. 94 - Opportunistic infections: management and prevention. In: Cohen, J., Powderly, W.G., Opal, S.M. (Eds.), Infectious Diseases, Fourth edition. Elsevier. https://doi.org/10.1016/B978-0-7020-6285-8.00094-0. 850-858.e1.

Dekaboruah, E., Suryavanshi, M.V., Chettri, D., Verma, A.K., 2020. Human microbiome: an academic update on human body site specific surveillance and its possible role. Arch. Microbiol. 202, 2147–2167. https://doi.org/10.1007/s00203-020-01931-x.

Deneke, C., Rentzsch, R., Renard, B.Y., 2017. PaPrBaG: a machine learning approach for the detection of novel pathogens from NGS data. Sci. Rep. 7, 39194. https://doi.org/10.1038/srep39194.

Dhariwal, A., Chong, J., Habib, S., King, I.L., Agellon, L.B., Xia, J., 2017. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res. 45, W180–W188. https://doi.org/10.1093/nar/gkx295.

Egli, A., Schrenzel, J., Greub, G., 2020. Digital microbiology. Clin. Microbiol. Infect. 26, 1324–1331. https://doi.org/10.1016/j.cmi.2020.06.023.

Estaki, M., Jiang, L., Bokulich, N.A., McDonald, D., González, A., Kosciolek, T., Martino, C., Zhu, Q., Birmingham, A., Vázquez-Baeza, Y., Dillon, M.R., Bolyen, E., Caporaso, J.G., Knight, R., 2020. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. Curr. Protoc. Bioinformatics 70. https://doi.org/10.1002/cpbi.100 e100.

Facciolà, A., Riso, R., Avventuroso, E., Visalli, G., Delia, S.A., Laganà, P., 2017. Campylobacter: from microbiology to prevention. J. Prev. Med. Hyg. 58, E79–E92.

Faergemann, J., 2002. Atopic dermatitis and fungi. Clin. Microbiol. Rev. 15, 545–563. https://doi.org/10.1128/CMR.15.4.545-563.2002.

Faron, M.L., Buchan, B.W., Coon, C., Liebregts, T., van Bree, A., Jansz, A.R., Soucy, G., Korver, J., Ledeboer, N.A., 2016a. Automatic digital analysis of chromogenic media for vancomycin-resistant-Enterococcus screens using Copan WASPLab. J. Clin. Microbiol. 54, 2464–2469. https://doi.org/10.1128/JCM.01040-16.

Faron, M.L., Buchan, B.W., Vismara, C., Lacchini, C., Bielli, A., Gesu, G., Liebregts, T., van Bree, A., Jansz, A., Soucy, G., Korver, J., Ledeboer, N.A., 2016b. Automated scoring of chromogenic media for detection of methicillin-resistant Staphylococcus aureus by use of WASPLab image analysis software. J. Clin. Microbiol. 54, 620–624. https://doi.org/10.1128/JCM.02778-15.

Feehily, C., Crosby, D., Walsh, C.J., Lawton, E.M., Higgins, S., McAuliffe, F.M., Cotter, P. D., 2020. Shotgun sequencing of the vaginal microbiome reveals both a species and functional potential signature of preterm birth. npj Biofilms Microbiomes 6, 1–9. https://doi.org/10.1038/s41522-020-00162-8.

Fettweis, J.M., Serrano, M.G., Brooks, J.P., Edwards, D.J., Girerd, P.H., Parikh, H.I., Huang, B., Arodz, T.J., Edupuganti, L., Glascock, A.L., Xu, J., Jimenez, N.R., Vivadelli, S.C., Fong, S.S., Sheth, N.U., Jean, S., Lee, V., Bokhari, Y.A., Lara, A.M., Mistry, S.D., Duckworth, R.A., Bradley, S.P., Koparde, V.N., Orenda, X.V., Milton, S. H., Rozycki, S.K., Matveyev, A.V., Wright, M.L., Huzurbazar, S.V., Jackson, E.M., Smirnova, E., Korlach, J., Tsai, Y.-C., Dickinson, M.R., Brooks, J.L., Drake, J.I., Chaffin, D.O., Sexton, A.L., Gravett, M.G., Rubens, C.E., Wijesooriya, N.R., Hendricks-Muñoz, K.D., Jefferson, K.K., Strauss, J.F., Buck, G.A., 2019. The vaginal microbiome and preterm birth. Nat. Med. 25, 1012–1021. https://doi.org/10.1038/s41591-019-0450-2.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J.T., Blum, M., Hutter, F., 2019. Auto-sklearn: efficient and robust automated machine learning. In: Hutter, F., Kotthoff, L., Vanschoren, J. (Eds.), Automated Machine Learning: Methods, Systems, Challenges. Springer International Publishing, Cham, pp. 113–134. https://doi.org/10.1007/978-3-030-05318-5_6. The Springer Series on Challenges in Machine Learning.

Florin, L., Maelegheer, K., Muyldermans, A., Van Esbroeck, M., Nulens, E., Emmerechts, J., 2018. Evaluation of the CellaVision DM96 advanced RBC application for screening and follow-up of malaria infection. Diagn. Microbiol. Infect. Dis. 90, 253–256. https://doi.org/10.1016/j.diagmicrobio.2017.12.002.

Fox, H., Purcell, R.H., Chanock, R.M., 1969. Characterization of a newly identified mycoplasma (Mycoplasma orale type 3) from the human oropharynx. J. Bacteriol. 98, 36–43.
Franasiak, J.M., Scott, R.T., 2015. Reproductive tract microbiome in assisted reproductive technologies. Fertil. Steril. 104, 1364–1371. https://doi.org/10.1016/j.fertnstert.2015.10.012.
Franco-Duarte, R., Černáková, L., Kadam, S., Kaushik, S., K, Salehi, B., Bevilacqua, A., Corbo, M.R., Antolak, H., Dybka-Stępień, K., Leszczewicz, M., Relison Tintino, S., Alexandrino de Souza, V.C., Sharifi-Rad, J., Melo Coutinho, H.D., Martins, N., Rodrigues, C.F., 2019. Advances in chemical and biological methods to identify microorganisms—from past to present. Microorganisms 7. https://doi.org/10.3390/microorganisms7050130.
Fukui, H., Nishida, A., Matsuda, S., Kira, F., Watanabe, S., Kuriyama, M., Kawakami, K., Aikawa, Y., Oda, N., Arai, K., Matsunaga, A., Nonaka, M., Nakai, K., Shinmura, W., Matsumoto, M., Morishita, S., Takeda, A.K., Miwa, H., 2020. Usefulness of machine learning-based gut microbiome analysis for identifying patients with irritable bowels syndrome. J. Clin. Med. 9, 2403. https://doi.org/10.3390/jcm9082403.
Ghannam, R.B., Techtmann, S.M., 2021. Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. Comput. Struct. Biotechnol. J. 19, 1092–1107. https://doi.org/10.1016/j.csbj.2021.01.028.
Grice, E.A., Kong, H.H., Conlan, S., Deming, C.B., Davis, J., Young, A.C., Bouffard, G.G., Blakesley, R.W., Murray, P.R., Green, E.D., Turner, M.L., Segre, J.A., 2009. Topographical and temporal diversity of the human skin microbiome. Science 324, 1190–1192. https://doi.org/10.1126/science.1171700.
Grogan, M.D., Bartow-McKenney, C., Flowers, L., Knight, S.A.B., Uberoi, A., Grice, E.A., 2019. Research techniques made simple: profiling the skin microbiota. J. Invest. Dermatol. 139 https://doi.org/10.1016/j.jid.2019.01.024, 747–752.e1.
Handelsman, J., 2004. Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68 https://doi.org/10.1128/MMBR.68.4.669-685.2004, 669 LP – 685.
Helal, Z.M., Rizk, D.E., Adel El-Sokkary, M.M., Hassan, R., 2020. Prevalence and characterization of Streptococcus pyogenes clinical isolates from different hospitals and clinics in Mansoura. Int. J. Microbiol. 2020 https://doi.org/10.1155/2020/5814945 e5814945.
Huson, D.H., 2015. MEtaGenome ANalyzer (MEGAN): metagenomic expert resource. In: Encyclopedia of Metagenomics. Springer US, pp. 383–389. https://doi.org/10.1007/978-1-4899-7478-5_4.
Huson, D.H., Albrecht, B., Bağci, C., Bessarab, I., Górska, A., Jolic, D., Williams, R.B.H., 2018. MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. Biol. Direct 13. https://doi.org/10.1186/s13062-018-0208-7.
Jasner, Y., Belogolovski, A., Ben-Itzhak, M., Koren, O., Louzoun, Y., 2021. Microbiome preprocessing machine learning pipeline. Front. Immunol. 12.
Ji, P., Zhang, Y., Wang, J., Zhao, F., 2017. MetaSort untangles metagenome assembly by reducing microbial community complexity. Nat. Commun. 8, 14306. https://doi.org/10.1038/ncomms14306.
Kan, B., Zhou, H., Du, P., Zhang, W., Lu, X., Qin, T., Xu, J., 2018. Transforming bacterial disease surveillance and investigation using whole-genome sequence to probe the trace. Front. Med. 12, 23–33. https://doi.org/10.1007/s11684-017-0607-7.
Kashyap, P.C., Chia, N., Nelson, H., Segal, E., Elinav, E., 2017. Microbiome at the frontier of personalized medicine. Mayo Clin. Proc. 92, 1855–1864. https://doi.org/10.1016/j.mayocp.2017.10.004.
Keegan, K.P., Glass, E.M., Meyer, F., 2016. MG-RAST, a metagenomics service for analysis of microbial community structure and function. In: Methods in Molecular Biology. Humana Press Inc., pp. 207–233. https://doi.org/10.1007/978-1-4939-3369-3_13
Kelly, D., Yang, L., Pei, Z., 2018. Gut microbiota, fusobacteria, and colorectal cancer. Diseases 6. https://doi.org/10.3390/diseases6040109.
Khan, R., Petersen, F.C., Shekhar, S., 2019. Commensal bacteria: an emerging player in defense against respiratory pathogens. Front. Immunol. 10 https://doi.org/10.3389/fimmu.2019.01203.
Khatoon, Z., McTiernan, C.D., Suuronen, E.J., Mah, T.-F., Alarcon, E.I., 2018. Bacterial biofilm formation on implantable devices and approaches to its treatment and prevention. Heliyon 4. https://doi.org/10.1016/j.heliyon.2018.e01067.
Kong, H.H., 2011. Skin microbiome: genomics-based insights into the diversity and role of skin microbes. Trends Mol. Med. 17, 320–328. https://doi.org/10.1016/j.molmed.2011.01.013.
Koohi-Moghadam, M., Borad, M.J., Tran, N.L., Swanson, K.R., Boardman, L.A., Sun, H., Wang, J., 2019. MetaMarker: a pipeline for de novo discovery of novel metagenomic biomarkers. Bioinformatics 35, 3812–3814. https://doi.org/10.1093/bioinformatics/btz123.
Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D., Charles, T.C., 2015. Current and future resources for functional metagenomics. Front. Microbiol. 6, 1196. https://doi.org/10.3389/fmicb.2015.01196.
Larsen, B., Monif, G.R.G., 2001. Understanding the bacterial flora of the female genital tract. Clin. Infect. Dis. 32, e69–e77. https://doi.org/10.1086/318710.
Leger, A.J.St, Desai, J.V., Drummond, R.A., Kugadas, A., Almaghrabi, F., Silver, P., Raychaudhuri, K., Gadjeva, M., Iwakura, Y., Lionakis, M.S., Caspi, R.R., 2017. An ocular commensal protects against corneal infection by driving an interleukin 17 response from mucosal γδ T cells. Immunity 47. https://doi.org/10.1016/j.immuni.2017.06.014, 148–158.e5.
Li, J., Qian, J.-M., 2020. Artificial intelligence in inflammatory bowel disease: current status and opportunities. Chin. Med. J. 133, 757–759. https://doi.org/10.1097/CM9.0000000000000714.
Li, F., Chen, C., Wei, W., Wang, Z., Dai, J., Hao, L., Song, L., Zhang, X., Zeng, L., Du, H., Tang, H., Liu, N., Yang, H., Wang, J., Madsen, L., Brix, S., Kristiansen, K., Xu, X., Li, J., Wu, R., Jia, H., 2018. The metagenome of the female upper reproductive tract. Gigascience 7. https://doi.org/10.1093/gigascience/giy107.
Liu, H., Whitehouse, C.A., Li, B., 2018. Presence and persistence of Salmonella in water: the impact on microbial quality of water and food safety. Front. Public Health 6. https://doi.org/10.3389/fpubh.2018.00159.
Lo, C., Marculescu, R., 2019. MetaNN: accurate classification of host phenotypes from metagenomic data using neural networks. BMC Bioinforma. 20, 314. https://doi.org/10.1186/s12859-019-2833-2.
Lupolova, N., Dallman, T.J., Holden, N.J., Gally, D.L., 2017. Patchy promiscuity: machine learning applied to predict the host specificity of Salmonella enterica and Escherichia coli. Microb. Genom. 3 https://doi.org/10.1099/mgen.0.000135.
Macfarlane, G.T., Cummings, J.H., 1999. Probiotics and prebiotics: can regulating the activities of intestinal bacteria benefit health? BMJ 318, 999–1003.
Magne, F., Gotteland, M., Gauthier, L., Zazueta, A., Pesoa, S., Navarrete, P., Balamurugan, R., 2020. The firmicutes/bacteroidetes ratio: a relevant marker of gut dysbiosis in obese patients? Nutrients 12. https://doi.org/10.3390/nu12051474.
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., Shigdel, R., Stres, B., Vilne, B., Yousef, M., Zdravevski, E., Tsamardinos, I., de Santa, Carrillo, Pau, E., Claesson, M.J., Moreno-Indias, I., Truu, J., 2021. Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. Front. Microbiol. 12, 634511. https://doi.org/10.3389/fmicb.2021.634511.
Markowitz, V.M., Chen, I.M.A., Chu, K., Szeto, E., Palaniappan, K., Jacob, B., Ratner, A., Liolios, K., Pagani, I., Huntemann, M., Mavromatis, K., Ivanova, N.N., Kyrpides, N.C., 2012. IMG/M-HMP: a metagenome comparative analysis system for the human microbiome project. PLoS One 7. https://doi.org/10.1371/journal.pone.0040151.
Mashima, I., Theodorea, C.F., Thaweboon, B., Thaweboon, S., Nakazawa, F., 2016. Identification of Veillonella species in the tongue biofilm by using a novel one-step polymerase chain reaction method. PLoS One 11. https://doi.org/10.1371/journal.pone.0157516.
McDonald, D., Hyde, E., Debelius, J.W., Morton, J.T., Gonzalez, A., Ackermann, G., Aksenov, A.A., Behsaz, B., Brennan, C., Chen, Y., DeRight Goldasich, L., Dorrestein, P.C., Dunn, R.R., Fahimipour, A.K., Gaffney, J., Gilbert, J.A., Gogul, G., Green, J.L., Hugenholtz, P., Humphrey, G., Huttenhower, C., Jackson, M.A., Janssen, S., Jeste, D.V., Jiang, L., Kelley, S.T., Knights, D., Kosciolek, T., Ladau, J., Leach, J., Marotz, C., Meleshko, D., Melnik, A.V., Metcalf, J.L., Mohimani, H., Montassier, E., Navas-Molina, J., Nguyen, T.T., Peddada, S., Pevzner, P., Pollard, K. S., Rahnavard, G., Robbins-Pianka, A., Sangwan, N., Shorenstein, J., Smarr, L., Song, S.J., Spector, T., Swafford, A.D., Thackray, V.G., Thompson, L.R., Tripathi, A., Vázquez-Baeza, Y., Vrbanac, A., Wischmeyer, P., Wolfe, E., Zhu, Q., Consortium, American Gut, Knight, R., 2018. American gut: an open platform for citizen science microbiome research. mSystems 3. https://doi.org/10.1128/mSystems.00031-18 e00031-18.
Meena, K.K., Sorty, A.M., Bitla, U.M., Choudhary, K., Gupta, P., Pareek, A., Singh, D.P., Prabha, R., Sahu, P.K., Gupta, V.K., Singh, H.B., Krishanani, K.K., Minhas, P.S., 2017. Abiotic stress responses and microbe-mediated mitigation in plants: the omics strategies. Front. Plant Sci. 8 https://doi.org/10.3389/fpls.2017.00172.
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 Bioinforma. 9, 386. https://doi.org/10.1186/1471-2105-9-386.
Najar, I.N., Sherpa, M.T., Das, Sayak, Das, Saurav, Thakur, N., 2020. Diversity analysis and metagenomic insights into antibiotic and metal resistance among Himalayan hot spring bacteriobiome insinuating inherent environmental baseline levels of antibiotic and metal tolerance. J. Glob. Antimicrob. Resist. 21, 342–352. https://doi.org/10.1016/j.jgar.2020.03.026.
National Research Council (US) Committee, 2000. Health Applications of the Internet. Networking Health: Prescriptions for the Internet. National Academies Press (US).
Neu, J., Rushing, J., 2011. Cesarean versus vaginal delivery: long term infant outcomes and the hygiene hypothesis. Clin. Perinatol. 38, 321–331. https://doi.org/10.1016/j.clp.2011.03.008.
Oh, M., Zhang, L., 2020. DeepMicro: deep representation learning for disease prediction based on microbiome data. Sci. Rep. 10, 6026. https://doi.org/10.1038/s41598-020-63159-5.
Okada, F., Ando, Y., Nakayama, T., Tanoue, S., Ishii, R., Ono, A., Watanabe, M., Takaki, H., Maeda, T., Mori, H., 2011. Pulmonary thin-section CT findings in acute Moraxella catarrhalis pulmonary infection. Br. J. Radiol. 84, 1109–1114. https://doi.org/10.1259/bjr/42762966.
Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G.A., Papanikolaou, N., Kotoulas, G., Arvanitidis, C., Iliopoulos, I., 2015. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. Bioinforma. Biol. Insights 9, 75–88. https://doi.org/10.4137/BBI.S12462.
Palma, S.I.C.J., Traguedo, A.P., Porteira, A.R., Frias, M.J., Gamboa, H., Roque, A.C.A., 2018. Machine learning for the meta-analyses of microbial pathogens' volatile signatures. Sci. Rep. 8, 3360. https://doi.org/10.1038/s41598-018-21544-1.
Pasolli, E., Truong, D.T., Malik, F., Waldron, L., Segata, N., 2016. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. PLoS Comput. Biol. 12 https://doi.org/10.1371/journal.pcbi.1004977 e1004977.
Pattnaik, S., Mohapatra, B., Gupta, A., 2021. Plant growth-promoting microbe mediated uptake of essential nutrients (Fe, P, K) for crop stress management: microbe–soil–plant continuum. Front. Agron. 3.

Pearman, W.S., Freed, N.E., Silander, O.K., 2020. Testing the advantages and disadvantages of short- and long- read eukaryotic metagenomics using simulated reads. BMC Bioinforma. 21, 220. https://doi.org/10.1186/s12859-020-3528-4.
Pereira, F.C., Berry, D., 2017. Microbial nutrient niches in the gut. Environ. Microbiol. 19, 1366–1378. https://doi.org/10.1111/1462-2920.13659.
Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J.A., Bonazzi, V., McEwen, J.E., Wetterstrand, K.A., Deal, C., Baker, C.C., Di Francesco, V., Howcroft, T.K., Karp, R.W., Lunsford, R.D., Wellington, C.R., Belachew, T., Wright, M., Giblin, C., David, H., Mills, M., Salomon, R., Mullins, C., Akolkar, B., Begg, L., Davis, C., Grandison, L., Humble, M., Khalsa, J., Little, A.R., Peavy, H., Pontzer, C., Portnoy, M., Sayre, M.H., Starke-Reed, P., Zakhari, S., Read, J., Watson, B., Guyer, M., 2009. The NIH human microbiome project. Genome Res. 19, 2317–2323. https://doi.org/10.1101/gr.096651.109.
Pickard, J.M., Zeng, M.Y., Caruso, R., Núñez, G., 2017. Gut microbiota: role in pathogen colonization, immune responses and inflammatory disease. Immunol. Rev. 279, 70–89. https://doi.org/10.1111/imr.12567.
Plummer, E.L., Vodstrcil, L.A., Bodiyabadu, K., Murray, G.L., Doyle, M., Latimer, R.L., Fairley, C.K., Payne, M., Chow, E.P.F., Garland, S.M., Bradshaw, C.S., 2021. Are Mycoplasma hominis, Ureaplasma urealyticum and Ureaplasma parvum associated with specific genital symptoms and clinical signs in non-pregnant women? Clin. Infect. Dis. https://doi.org/10.1093/cid/ciab061.
Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., Konstantinidis, K.T., 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 9. https://doi.org/10.1371/journal.pone.0093827 e93827.
Poutahidis, T., Springer, A., Levkovich, T., Qi, P., Varian, B.J., Lakritz, J.R., Ibrahim, Y. M., Chatzigiagkos, A., Alm, E.J., Erdman, S.E., 2014. Probiotic microbes sustain youthful serum testosterone levels and testicular size in aging mice. PLoS One 9. https://doi.org/10.1371/journal.pone.0084877.
Puchenkova, S.G., 1996. Enterobacteria in areas of water along the Crimean coast. Mikrobiol. Z. 58, 3–7.
Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, Shaochuan, Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, Shengting, Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, Songgang, Qin, N., Yang, H., Wang, Jian, Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S.D., Wang, Jun, 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65. https://doi.org/10.1038/nature08821.
Qiu, H., KuoLee, R., Harris, G., Zhou, H., Miller, H., Patel, G.B., Chen, W., 2011. Acinetobacter baumannii infection inhibits airway eosinophilia and lung pathology in a mouse model of allergic asthma. PLoS One 6. https://doi.org/10.1371/journal.pone.0022004.
Qu, K., Guo, F., Liu, X., Lin, Y., Zou, Q., 2019. Application of machine learning in microbiology. Front. Microbiol. 10 https://doi.org/10.3389/fmicb.2019.00827.
Reiman, D., Metwally, A., Sun, J., Dai, Y., 2021. Meta-Signer: Metagenomic Signature Identifier Based Onrank Aggregation of Features. https://doi.org/10.12688/f1000research.27384.1.
Rhoads, D.D., 2020. Computer vision and artificial intelligence are emerging diagnostic tools for the clinical microbiologist. J. Clin. Microbiol. 58 https://doi.org/10.1128/JCM.00511-20.
Richards, A.L., Muehlbauer, A.L., Alazizi, A., Burns, M.B., Findley, A., Messina, F., Gould, T.J., Cascardo, C., Pique-Regi, R., Blekhman, R., Luca, F., 2019. Gut microbiota has a widespread and modifiable effect on host gene regulation. mSystems 4. https://doi.org/10.1128/mSystems.00323-18 e00323-18.
Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., Tuohy, K., 2018. Gut microbiota functions: metabolism of nutrients and other food components. Eur. J. Nutr. 57, 1–24. https://doi.org/10.1007/s00394-017-1445-8.
Schmid, M., Muri, J., Melidis, D., Varadarajan, A.R., Somerville, V., Wicki, A., Moser, A., Bourqui, M., Wenzel, C., Eugster-Meier, E., Frey, J.E., Irmler, S., Ahrens, C.H., 2018. Comparative genomics of completely sequenced *Lactobacillus helveticus* genomes provides insights into strain-specific genes and resolves metagenomics data down to the strain level. Front. Microbiol. 9.
Schmidt, J., Marques, M.R.G., Botti, S., Marques, M.A.L., 2019. Recent advances and applications of machine learning in solid-state materials science. npj Comput. Mater. 5, 1–36. https://doi.org/10.1038/s41524-019-0221-0.
Schnabl, B., 2014. The microbiome and the liver. Gastroenterol. Hepatol. (N Y) 10, 519–521.
Schwarze, K., Buchanan, J., Fermont, J.M., Dreau, H., Tilley, M.W., Taylor, J.M., Antoniou, P., Knight, S.J.L., Camps, C., Pentony, M.M., Kvikstad, E.M., Harris, S., Popitsch, N., Pagnamenta, A.T., Schuh, A., Taylor, J.C., Wordsworth, S., 2020. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. Genet. Med. 22, 85–94. https://doi.org/10.1038/s41436-019-0618-7.
Shamsaddini, A., Dadkhah, K., Gillevet, P.M., 2020. BiomMiner: an advanced exploratory microbiome analysis and visualization pipeline. PLoS One 15. https://doi.org/10.1371/journal.pone.0234860 e0234860.
Siddesh, G.M., Krutika, S., Srinivasa, K.G., Siddiqui, N., 2021. Healthcare data analytics using artificial intelligence. In: Srinivasa, K.G., G. M., S, Sekhar, S.R.M. (Eds.), Artificial Intelligence for Information Management: A Healthcare Perspective, Studies in Big Data. Springer, Singapore, pp. 45–85.
Slade, H.D., Vetter, J.K., 1956. Studies on Streptococcus pyogenes I. J. Bacteriol. 71, 236–243.
Sreekumar, S., Baer, Z.C., Pazhamalai, A., Gunbas, G., Grippo, A., Blanch, H.W., Clark, D. S., Toste, F.D., 2015. Production of an acetone-butanol-ethanol mixture from Clostridium acetobutylicum and its conversion to high-value biofuels. Nat. Protoc. 10, 528–537. https://doi.org/10.1038/nprot.2015.029.
Stefan, L., Hartmann, M., Engbersen, N., Six, J., Schöb, C., 2021. Positive effects of crop diversity on productivity driven by changes in soil microbial composition. Front. Microbiol. 12.
Straif-Bourgeois, S., Ratard, R., Kretzschmar, M., 2014. Infectious disease epidemiology. In: Handbook of Epidemiology, pp. 2041–2119. https://doi.org/10.1007/978-0-387-09834-0_34.
Sudhakar, P., Machiels, K., Verstockt, B., Korcsmaros, T., Vermeire, S., 2021. Computational biology and machine learning approaches to understand mechanistic microbiome-host interactions. Front. Microbiol. 12 https://doi.org/10.3389/fmicb.2021.618856.
Talamantes-Becerra, B., Carling, J., Kilian, A., Georges, A., 2020. Discovery of thermophilic Bacillales using reduced-representation genotyping for identification. BMC Microbiol. 20, 114. https://doi.org/10.1186/s12866-020-01800-z.
Tap, J., Derrien, M., Törnblom, H., Brazeilles, R., Cools-Portier, S., Doré, J., Störsrud, S., Le Nevé, B., Öhman, L., Simrén, M., 2017. Identification of an intestinal microbiota signature associated with severity of irritable bowel syndrome. Gastroenterology 152. https://doi.org/10.1053/j.gastro.2016.09.049, 111–123.e8.
Tett, A., Pasolli, E., Masetti, G., Ercolini, D., Segata, N., 2021. Prevotella diversity, niches and interactions with the human host. Nat. Rev. Microbiol. 1–15. https://doi.org/10.1038/s41579-021-00559-y.
Thursby, E., Juge, N., 2017. Introduction to the human gut microbiota. Biochem. J. 474, 1823–1836. https://doi.org/10.1042/BCJ20160510.
Tong, S.Y.C., Davis, J.S., Eichenberger, E., Holland, T.L., Fowler, V.G., 2015. Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management. Clin. Microbiol. Rev. 28, 603–661. https://doi.org/10.1128/CMR.00134-14.
Topalcengiz, Z., Danyluk, M.D., 2019. Fate of generic and Shiga toxin-producing Escherichia coli (STEC) in Central Florida surface waters and evaluation of EPA worst case water as standard medium. Food Res. Int. 120, 322–329. https://doi.org/10.1016/j.foodres.2019.02.045.
Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., Gordon, J.I., 2009. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci. Transl. Med. 1 https://doi.org/10.1126/scitranslmed.3000322, 6ra14.
Ursell, L.K., Metcalf, J.L., Parfrey, L.W., Knight, R., 2012. Defining the human microbiome. Nutr. Rev. 70, S38–S44. https://doi.org/10.1111/j.1753-4887.2012.00493.x.
Valenti, P., Rosa, L., Capobianco, D., Lepanto, M.S., Schiavi, E., Cutone, A., Paesano, R., Mastromarino, P., 2018. Role of lactobacilli and lactoferrin in the mucosal cervicovaginal defense. Front. Immunol. 9 https://doi.org/10.3389/fimmu.2018.00376.
Wang, W.-L., Xu, S.-Y., Ren, Z.-G., Tao, L., Jiang, J.-W., Zheng, S.-S., 2015. Application of metagenomics in the human gut microbiome. World J. Gastroenterol. 21, 803–814. https://doi.org/10.3748/wjg.v21.i3.803.
Wang, H., Lee, I.-S., Braun, C., Enck, P., 2016. Effect of probiotics on central nervous system functions in animals and humans: a systematic review. J Neurogastroenterol. Motil. 22, 589–605. https://doi.org/10.5056/jnm16018.
Wani, A.K., Akhtar, N., Datta, B., Pandey, J., Amin-ul Mannan, M., 2021. Chapter 14 - Cyanobacteria-derived small molecules: a new class of drugs. In: Kumar, A., Singh, J., Samuel, J. (Eds.), Volatiles and Metabolites of Microbes. Academic Press, pp. 283–303. https://doi.org/10.1016/B978-0-12-824523-1.00003-1.
Wani, Atif Khurshid, Akhtar, N., Sher, F., Navarrete, A.A., Américo-Pinheiro, J.H.P., 2022a. Microbial adaptation to different environmental conditions: molecular perspective of evolved genetic and cellular systems. Arch. Microbiol. 204, 144. https://doi.org/10.1007/s00203-022-02757-5.
Wani, A.K., Rahayu, F., Kadarwati, F.T., Suhara, C., Singh, R., Dhanjal, D.S., Akhtar, N., Mir, T.G., Chopra, C., 2022b. Metagenomic screening strategies for bioprospecting enzymes from environmental samples. IOP Conf. Ser. Earth Environ. Sci. 974 https://doi.org/10.1088/1755-1315/974/1/012003, 012003.
Watanabe, N., Bandoh, S., Ishii, T., Negayama, K., Kadowaki, N., Yokota, K., 2017. Infective endocarditis and infected aneurysm caused by Streptococcus dysgalactiae subsp. equisimilis: a case report. Clin. Case Rep. 5, 187–192. https://doi.org/10.1002/ccr3.829.
Whalen, J.G., Mully, W.T., English, C.J., 2007. Spontaneous Citrobacter freundii infection in an immunocompetent patient. Arch. Dermatol. 143, 124–125. https://doi.org/10.1001/archderm.143.1.124.
Wong, D., Nielsen, T.B., Bonomo, R.A., Pantapalangkoor, P., Luna, B., Spellberg, B., 2017. Clinical and pathophysiological overview of Acinetobacter infections: a century of challenges. Clin. Microbiol. Rev. 30, 409–447. https://doi.org/10.1128/CMR.00058-16.
Wu, H., Cai, L., Li, D., Wang, X., Zhao, S., Zou, F., Zhou, K., 2018. Metagenomics biomarkers selected for prediction of three different diseases in Chinese population. Biomed. Res. Int. 2018, 2936257. https://doi.org/10.1155/2018/2936257.
Yang, F., Zou, Q., 2020. mAML: An Automated Machine Learning Pipeline with a Microbiome Repository for Human Disease Classification. Database (Oxford) 2020, baaa050. https://doi.org/10.1093/database/baaa050.
Zhang, S., Li, S., Gu, W., den Bakker, H., Boxrud, D., Taylor, A., Roe, C., Driebe, E., Engelthaler, D.M., Allard, M., Brown, E., McDermott, P., Zhao, S., Bruce, B.B., Trees, E., Fields, P.I., Deng, X., 2019. Zoonotic source attribution of Salmonella enterica serotype typhimurium using genomic surveillance data, United States. Emerg. Infect. Dis. 25, 82–91. https://doi.org/10.3201/eid2501.180835.

Zhong, C., Chen, C., Wang, L., Ning, K., 2021. Integrating pan-genome with metagenome for microbial community profiling. Comput. Struct. Biotechnol. J. 19, 1458–1466. https://doi.org/10.1016/j.csbj.2021.02.021.

Zhou, Y.-H., Gallins, P., 2019. A review and tutorial of machine learning methods for microbiome host trait prediction. Front. Genet. 10, 579. https://doi.org/10.3389/fgene.2019.00579.

Zhou, G., Jiang, J.-Y., Ju, C.J.-T., Wang, W., 2019. Prediction of microbial communities for urban metagenomics using neural network approach. Hum. Genom. 13 https://doi.org/10.1186/s40246-019-0224-4.