Computational and Structural Biotechnology Journal 21 (2023) 2502–2513

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Software/Web server Article

CDEMI: Characterizing differences in microbial composition and function in microbiome data

Lidan Wang [a,b,1], Xiao Liang [a,1], Hao Chen [a], Lijie Cao [a], Lan Liu [a], Feng Zhu [d], Yubin Ding [b,*], Jing Tang [a,c,**], Youlong Xie [c,*]

[a] School of Basic Medicine, Chongqing Medical University, Chongqing 400016, China
[b] Department of Obstetrics and Gynecology, Women and Children's Hospital of Chongqing Medical University, Chongqing 401147, China
[c] Joint International Research Laboratory of Reproductive and Development, Department Reproductive Biology, School of Public Health, Chongqing Medical University, Chongqing 400016, China
[d] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 28 December 2022
Received in revised form 23 March 2023
Accepted 24 March 2023
Available online 25 March 2023

Keywords:
Microbiome
Microbial composition
Functional characterization
Microbial association
Metabolic pathway

ABSTRACT

Microbial communities influence host phenotypes through microbiota-derived metabolites and interactions between exogenous active substances (EASs) and the microbiota. Owing to the high dynamics of microbial community composition and difficulty in microbial functional analysis, the identification of mechanistic links between individual microbes and host phenotypes is complex. Thus, it is important to characterize variations in microbial composition across various conditions (for example, topographical locations, times, physiological and pathological conditions, and populations of different ethnicities) in microbiome studies. However, no web server is currently available to facilitate such characterization. Moreover, accurately annotating the functions of microbes and investigating the possible factors that shape microbial function are critical for discovering links between microbes and host phenotypes. Herein, an online tool, CDEMI, is introduced to discover microbial composition variations across different conditions, and five types of microbe libraries are provided to comprehensively characterize the functionality of microbes from different perspectives. These collective microbe libraries include (1) microbial functional pathways, (2) disease associations with microbes, (3) EASs associations with microbes, (4) bioactive microbial metabolites, and (5) human body habitats. In summary, CDEMI is unique in that it can reveal microbial patterns in distributions/compositions across different conditions and facilitate biological interpretations based on diverse microbe libraries. CDEMI is accessible at http://rdblab.cn/cdemi/

1. Introduction

Diverse microbial communities are inhabited by various body sites of humans [1]. These microbes influence host phenotypes [2,3] through microbiota-derived metabolites [4,5] and microbial-secreted proteins and cell compositions [6,7]. Moreover, some exogenous active substances (EASs) such as diet [8], drug metabolites [9], traditional medicines [10], and environmental toxins [11] have previously been reported to contribute to the variation of microbial communities [12], thereby influencing host phenotypes. Therefore, associations between microbial communities and various host phenotypes have attracted considerable attention in current metagenomic studies [13].

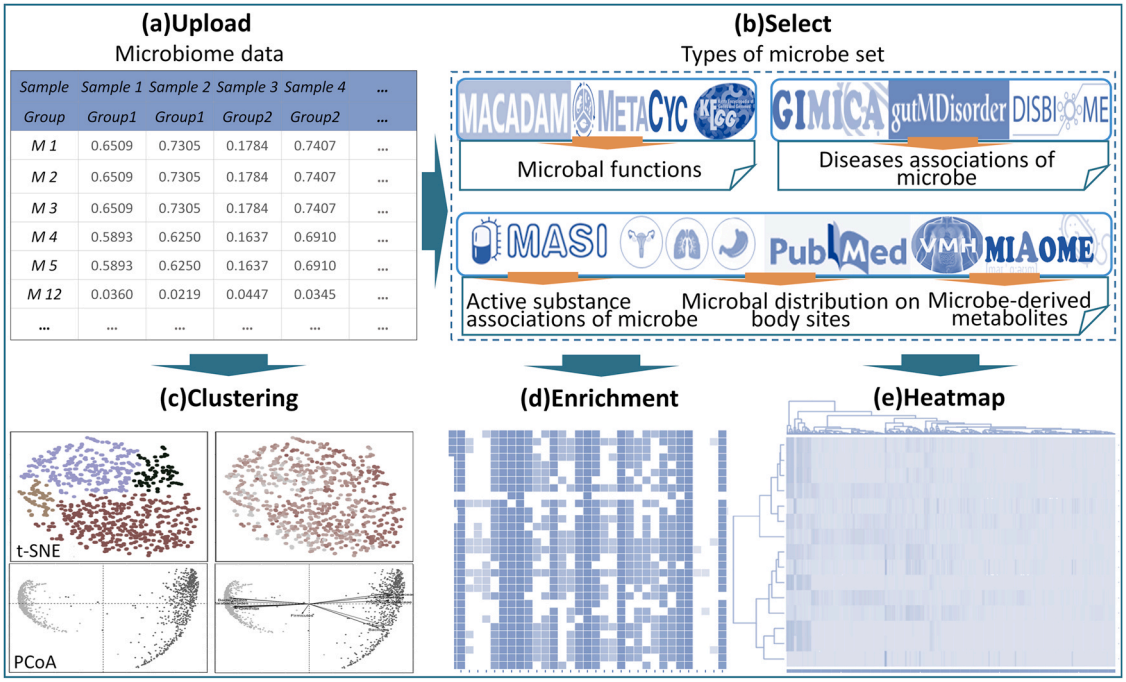
However, the high dynamics of human microbial composition and difficulty in functional analysis, have hindered efforts to define

Abbreviations: EASs, exogenous active substances; t-SNE, t-distributed Stochastic Neighbor Embedding; PCoA, Principal Coordinates Analysis; MDMs, microbial-derived metabolites; PERMANOVA, permutational multivariate analysis of variance; MSEA, microbe set enrichment analysis; LP, leprosy patients; HC, healthy control; OTUs, Operational Taxonomic Units; DR, diabetic retinopathy; eOTUs, enriched OTUs; dOTUs, depleted OTUs; oOTUs, ordinary OTUs; LEfSe, linear discriminant analysis Effect Size.

* Corresponding authors.
** Corresponding author at: School of Basic Medicine, Chongqing Medical University, Chongqing 400016, China.
E-mail addresses: dingyb@cqmu.edu.cn (Y. Ding), tang_jing@cqmu.edu.cn (J. Tang), 191029@cqmu.edu.cn (Y. Xie).

[1] The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

https://doi.org/10.1016/j.csbj.2023.03.044



Sample	Sample 1	Sample 2	Sample 3	Sample 4	...
Group	Group1	Group1	Group2	Group2	...
M 1	0.6509	0.7305	0.1784	0.7407	...
M 2	0.6509	0.7305	0.1784	0.7407	...
M 3	0.6509	0.7305	0.1784	0.7407	...
M 4	0.5893	0.6250	0.1637	0.6910	...
M 5	0.5893	0.6250	0.1637	0.6910	...
M 12	0.0360	0.0219	0.0447	0.0345	...
...	...	...	...	...	...

Fig. 1. The standard workflow of CDEMI: (a) Uploading metagenomics/16S rRNA gene sequencing data (microbiome abundance table); (b) Selection of the types of microbe libraries; (c) Characterization and visualization of the differences in microbial compositions across various conditions; (d) Annotation and enrichment of microbes from different perspectives; (e) Heatmap of enrichment analysis of the microbe set across all microbial samples.

the mechanistic links between individual microbes and host phenotypes [2]. These microbial community compositions are highly variable depending on topographical locations, times, and physiological and pathological conditions [14–16]. At the spatial scale, microbes from various physical niches (for example, spatially distinct habitats) can greatly differ in community compositions [17,18], at the temporal scale, microbial diversity and composition can change substantially across different stages of host development [19,20]. Studying variations in microbial community compositions under different conditions enables the development of an association between microbes and the host phenotype [21,22]. However, the active level of microbial function is also substantially diversified under different conditions [23–25]. Thus, it is also necessary to construct an online tool for characterizing variations in microbial composition and function under various conditions in microbiome studies.

Various powerful online tools have been designed to facilitate the identification of differential microbes under different conditions and microbial function annotations [26–28]. Some tools have been dedicated to linking microbial communities and biological functions, such as KEGG [29], MACADAM [30], HPMCD [31], and fusionDB [32]. Others, including DAnIEL [28], EasyMap [33], VITCOMIC [34], and PM2RA [35] were developed for the differential abundance analysis of microbes. Currently, only MicrobiomeAnalyst [36] and FunGeCo [37] support both, but neither identifies the phenotype-specific microbes nor displays substantial variations among different phenotypes. Moreover, none of the available online tools can visualize the microbial abundance distributions among different clusters of microbial samples for a given microbe and identify the EASs interacting with the microbiota (for example, herbal products). Thus, an online tool that can facilitate such characterization is urgently required. However, no such tool has been developed.

In this study, a novel web –server, CDEMI, was constructed. As illustrated in Fig. 1, CDEMI is unique in its capability of (**1**) discovering phenotype-specific microbes/subtype-specific microbes and visualizing substantial variations in distributions/compositions across various phenotypes/subtypes, (**2**) enabling us to gain functional insights from five different microbe libraries (integrating microbial functional pathways, disease associations with the microbes, EAS associations with the microbes, bioactive microbial metabolites, and human body habitats), and (**3**) offering enrichment analysis based on diverse microbial libraries and interactive visualization of the result. Collectively, CDEMI is distinguished for its capacity to characterize the differences in microbial community composition, to investigate the functions of microbes, and to discover the EAS interacting with microbiota, and is therefore expected to emerge as an indispensable complement to other available tools. CDEMI web-server can be freely accessible (without login requirement) at: http://rdblab.cn/cdemi/.

2. Methods

2.1. Benchmark datasets collected and analyzed in this study

To comprehensively illustrate the performance of CDEMI in analyzing phenotype-specific microbes, characterizing microbial function, and detecting the factors influencing microbes, the PubMed database was screened by searching keywords including 'Microbiome', 'Metagenomic', and '16 S rRNA' resulting in datasets from four publications. For evaluating the features of CDEMI in identifying phenotype-specific microbes, three benchmark datasets were included: microbiome data from two different human body sites (gut and nare) [38], skin microbiome data from two geographical locations (Hyderabad and Miraj in India) [39], and skin microbiome data from two physiological conditions (leprosy patients and healthy) [39]. To evaluate the features of CDEMI in characterizing microbial function, the fecal microbiome data of diabetic retinopathy and healthy samples from Shivaji et al. was used [40]. For evaluating the features of CDEMI in characterizing the exogenous substance influencing microbial functions, the salivary microbiome data of acute otitis media samples before and after amoxicillin treatment from Schrenzel et al. was used [41]. Detailed information on these benchmark data sets is presented in Table 1.

Table 1
Benchmark microbiome datasets involving various biological contexts (BC) were employed for illustrating the capacity of CDEMI from differential perspectives.

Author of publication	Sequence technology Sampling body site (BS)	The description of the sample, microbe, and study condition
Snyder et al.	16S rRNA sequencing BC: body sites BS: Gut & Nare	96 gut taxa and 80 nare taxa from 666 samples of 95 healthy individuals and individuals with prediabetes
Mande et al.	16S rRNA sequencing BC: living cities BS: Skin swab	skin microbiome of 25 phyla and 1016 genera from 88 samples of 30 healthy control and 58 individuals with leprosy patients from Hyderabad and Miraj in India
Shivaji et al.	16S rRNA sequencing BC: biological conditions BS: Feces	3539 OTUs from 58 samples taken from 30 healthy control and 28 samples from diabetic retinopathy
Schrenzel et al.	16S rRNA sequencing BC: biological conditions BS: Saliva	1656 OTUs from 36 acute otitis media saliva samples including 18 pre-amoxicillin and 18 post-amoxicillin samples

2.2. Required data formats of CDEMI input files and server implementation details

The standard format accepted by CDEMI is 'CSV' format with the dimension of m × n (m and n indicate the numbers of microbe (for example, taxon, species, genus, strains level) and microbial samples). The first two rows represent the sample ID and the sample groups. The sample ID is uniquely assigned according to user' preferences. For example, users could use 'sample 1', 'sample 2' to label the samples ID, and 'group 1', 'group 2' to label the samples group. Example files can be directly downloaded from the "Analysis" panel of CDEMI.

The CDEMI website was deployed on a server running Cent OS v7.0 operating system, Apache Tomcat servlet container, and Apache HTTP web server v2.4. Its interface was developed using R v4.1.2, and R package Shiny v0.13.1, running on Shiny-server v1.4.1.759. A variety of R software packages were used in the background processes, including shinythemes, shiny, shinyjs, shinyBS, shinydashboard, manhattanly, RColorBrewer, Seurat, and d3heatmap, CDEMI has been running smoothly for months and has been tested from various sites (such as Google Chrome, Mozilla Firefox, and Safari) worldwide and is freely accessible to all users without login requirements.

2.3. Characterizing differences in microbial composition across various conditions

Microbial communities play a fundamental role in human physiology, pathology, and behavioral phenotypes [42]. These microbial community compositions vary depending on body site, physiology, and environmental conditions [43,44]. Thus, it is important to characterize the differences in microbial community composition across various conditions in microbiome studies. In CDEMI, a nonlinear dimensionality-reduction algorithm, t-distributed Stochastic Neighbor Embedding (t-SNE) [45], is applied to visualize microbiome data. In particular, the microbial samples were clustered based on the similarity of microbial abundance, independent of the microbial collection site or time. Then, the distribution of microbes across various conditions or clusters was further determined using the 'featureplot' function in the 'Seurat' package for a specific microbe. Moreover, a Principal Coordinates Analysis (PCoA) of the Bray-Curtis distances was also performed to visualize the differences in microbial community composition [46], and Calculates scores (coordinates) [47] were used to select the significant microbes that correlated with the ordination ($|r| > 0.7$).

2.4. Annotating the microbe associations based on five types of libraries

Consider a group of microbes as a microbial set if there are established, empirically observed, or theoretically predicted functional associations among them. Based on these criteria, we constructed themed collections of microbial libraries. In CDEMI, five microbe libraries were constructed by integrating well-established databases including KEGG, MACADAM, VMH [48], MIAOME [49], GIMICA [50], gutMDisorder [51], Disbiome [52], and MASI [12], as well as manually searching the literature relevant to microbes. The information including microbe-derived metabolites and distribution in body site of microbiota were manually searched from PubMed [53] and Web of Science (https://clarivate.com/products/web-of-science/) databases using the combinations of keywords 'microbe', 'microbiota', 'microbiome', 'microbe', 'body site', 'habitat', 'metabolite', 'compound', 'molecule' to identify relevant studies published before November 2022. Publications should contain the following terms: the known microbial name and the information associated with the microbe (for example, bioactive microbial metabolite or human body habitat). To identify additional relevant publications, we checked the reference lists of the papers found in our search. Additional studies were included in our CDEMI if they contained (1) information on molecular compounds biosynthesized or metabolized by microbes, (2) information on disease associated with microbes, and (3) the body site distribution of microbes inhabiting humans.

2.4.1. Integrated data on microbial functions

To explore the function of microbes, we constructed a microbial function library in CDEMI by integrating microbial metabolic pathways from the KEGG and MACADAM databases. In sum, for the KEGG source, 427 metabolic pathways and 22,530 microbes (at species and genus level) were collected, for the MACADAM source, 1260 metabolic pathways and 3481 microbes (at strain, species, and genus level) were collected in CDEMI.

2.4.2. Integrated data on microbe-derived metabolites

Microbes produce a broad range of metabolic products that accumulate in high levels in the body. These metabolic products are regarded as microbial-derived metabolites (MDMs), which affect host health [54,55]. Exploring metabolites derived from microbes may provide mechanistic insights into the link between microbes and human diseases. Thus, CDEMI collectively contains MDMs by integrating well-established databases including VMH and MIAOME, as well as literature reviews. In sum, 11,898 links between 743 metabolites and 1489 microbes were included in CDEMI. We divided the 743 metabolites into 330 categories based on PubChem [56] and HMDB [57] databases.

2.4.3. Integrated data on disease association

Increasing research shows that a disturbed microbiome has been linked to hundreds of diseases, such as cancer, autoimmune, and cardiovascular diseases [58–60]. Thus, to gain a better understanding of the potential association between microbe and disease, CDEMI integrated the information of the known microbe-disease associations from the GIMICA, gutMDisorder, and Disbiome

Table 2
Statistics of data related to microbes in CDEMI and online databases (the first is the new tool proposed in this study), n: number, NA: not available.

Tools	Microbiome data analysis method	Microbe function (n)	Microbe-associated disease (n)	Microbe-derived metabolite (n)	Microbe-associated EASs (n)	Body sites of microbe inhabited (n)
CDEMI	t-SNE/ PCoA/ Enrichment analysis	1687	538	743	1296	84
KEGG	NA	427	NA	NA	NA	NA
MACADAM	NA	1260	NA	NA	NA	NA
GIMICA	NA	NA	192	NA	NA	9
gutMDisorder	NA	NA	123	NA	77	1 (gut)
Disbiome	NA	NA	375	NA	NA	50
MIAOME	NA	NA	NA	88	NA	NA
VMH	NA	NA	NA	126	NA	1 (gut)
MASI	NA	NA	56	NA	1296	NA

databases, and then constructed a microbe-associated disease library, which included 538 diseases associated with 1657 microbes.

2.4.4. Integrated data on EAS association

EASs, such as dietary, herbal, and environmental substances, which interact with host microbes, can affect host health and therapeutics [12]. To provide a better understanding of microbiota function, we constructed a microbe-related EASs library by integrating the data from 1296 EASs (divided into 46 categories) associated with 806 microbes from the MASI database.

2.4.5. Integrated data on the distribution of body habitats

Thousands of species coexist with the host and are distributed on the surfaces of the skin, intestine, and other mucous membranes [61]. These microbial distributions and compositions are highly variable depending on the physiological and pathological conditions [62]. Thus it is of great importance to characterize microbial composition differences across various body sites in microbiome studies. In CDEMI, we collected body sites inhabited by different microbes through literature reviews and constructed a body site habitats library. Consequently, CDEMI contains 942 microbes that habitat over 84 body sites.

In summary, CDEMI currently contains five microbe libraries involving 1687 microbial functions, 538 microbe-associated diseases, 743 microbe-derived metabolites, 1296 microbe-associated EASs, and 84 microbial body sites. These microbial associations provided by CDEMI are more diverse than those provided by the online databases included in our study (Table 2). Moreover, these online databases only provided microbe associations and were unable to analyze microbial composition variations across different conditions and support the microbe set enrichment analysis.

2.5. Microbe enrichment analysis based on five libraries types

Enrichment analysis can infer the collective functions of a set of microbes instead of a single microbe by identifying microbe sets sharing common attributes with the input microbe list. Based on the five microbe libraries in CDEMI, enrichment analysis was conducted to reveal the degree of aggregation of a functional role for the studied microbe list. Over-representation analysis using a hypergeometric test was applied for enrichment analyses. The statistical significance of enrichment was evaluated using a hypergeometric test with $p < 0.05$. Finally, an interactive Manhattan plot illustrating the enrichment results is displayed directly in CDEMI.

3. Results and discussion

3.1. Web service and operating procedure in CDEMI

To make the use of CDEMI convenient, the operating procedure implemented in this tool is provided in the following procedures (illustrated in Fig. 1). (**a**) uploading metagenomic/16 S rRNA gene sequencing data (microbiome abundance table); (**b**) selections of the types of microbe libraries (five libraries were provided for selection by users); (**c**) characterization and visualization of the differences in microbial compositions across various conditions by t-SNE analysis. t-SNE was applied for sample clustering, and Wilcoxon's rank-sum test was used to select the top distinguished microbes for each identified subcluster. Moreover, PCoA, as a visualization tool, has been widely used to visualize variations in microbial composition in current microbiome studies. In CDEMI, PCoA and permutational multivariate analysis of variance (PERMANOVA) for microbial community structures were performed [63], and scores (coordinates) were used to plot important species for PCoA results in CDEMI. (**d**) Microbe set enrichment analysis (MSEA) based on the annotation result of constructed microbe libraries; (**e**) heatmap of MSEA of the microbe set across all microbial samples. Detailed user manuals and website demos are provided in the '**Manual**' panel of CDEMI.

3.2. Characterizing the differences in microbe composition by CDEMI

3.2.1. Discovering the microbial composition differences across different body sites

The human body contains various habitats [64,65]. To evaluate the capacity of CDEMI to characterize the differences in microbial composition across body sites, the microbiome data from Snyder et al. [38] (listed in Table 1**)** were applied in this case study, which contained 96 gut taxa and 80 nare taxa from 666 samples (prediabetes and healthy). As shown in Fig. 2, there was an obvious difference in microbial composition between the gut and nare sites. Fig. 2A illustrates the relative abundance of dominant genera in the gut and nare microbiomes using a stack column plot. Interestingly, we found that the primary genus in the gut was *Bacteroide*s, but not in the nare. Conversely, the genus *Corynebacterium* in nare was primary, whereas it was almost absent in the gut. Similarly, the difference in microbial composition between the gut and nare sites was found via the distribution of the most abundant microbes (top ten microbes). As shown in Fig. 2B, for the gut, the relative abundance of microbes belonging to the phylum *Bacteroidetes* was approximately 46.1%, whereas the relative abundance of microbes belonging to the *Actinobacteria* phylum in the nare was approximately 48.2%.

Moreover, to examine the differences between gut and nare sites at the overall microbiome level, t-SNE clustering was applied based on all microbial samples. Fig. 2C reveals an obvious separation in the microbial community composition between the gut and nare groups. The abundance distributions of six representative microbes (discrepant between the gut and the nare) across all microbial samples are displayed in Fig. 2D. *Blautia* [66], *Faecalibacterium* [67], and *Bacteroides* [68] were specifically enriched in the gut, whereas *Propionibacterium* [69], *Corynebacterium* [70], and *Staphylococcus* [71] were specifically enriched in the nare. In addition, differences in microbial community composition between the gut and nare were visualized using PCoA of the Bray-Curtis distances (Fig. 2E). The

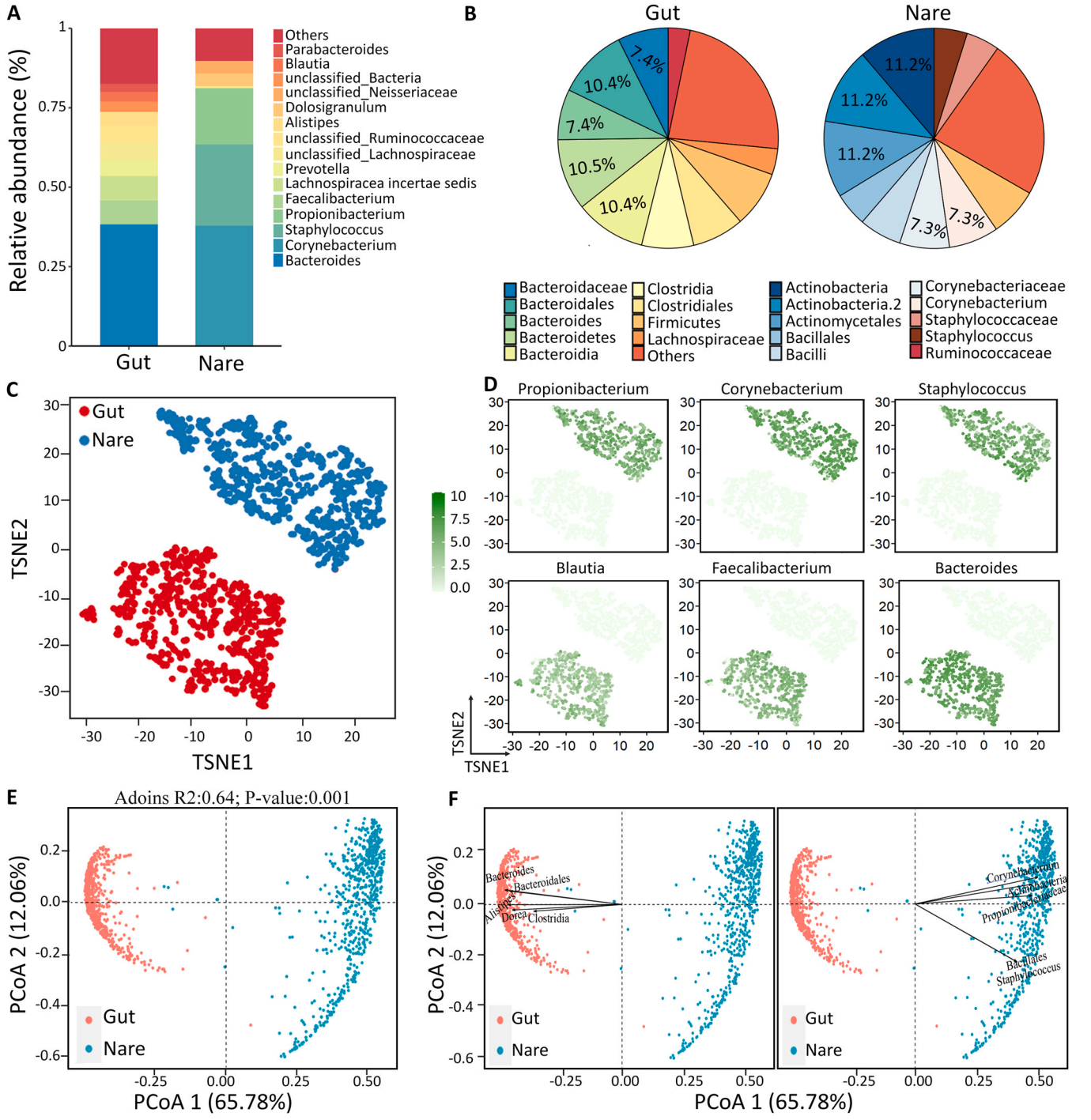


Fig. 2. Differences in microbial community composition between the gut and nare. Sample data was from Snyder et al. [38]. A. The relative abundance of the 15 dominant genera and others in the gut and nare microbiome regardless of healthy and prediabetes. The less abundant genera were grouped under "others". B. The relative abundance of the predominant microbes (top ten microbes) and others in the gut or nare. The less abundant microbes were grouped under "others". The relative abundance of dominant microbes was calculated using the mean relative abundance for each microbe of the gut and nare groups. C. The t-SNE plot of the gut and nare microbiome regardless of healthy and prediabetes. The colors represent the body sites (red: gut; blue: nare). D. The abundance distribution of representative microbes across all samples. The color key from light to dark indicates abundance levels from low to high. E. The differences in microbial community composition are shown by principal coordinates analysis (PCOA) of Bray-Curtis Distances, each symbol represents a sample. The color represents the body sites (red: gut; blue: nare). F. Biplot of PCoA with projected scores of major microbes which contributed to differences between the gut and nare sites.

results showed that the microbial composition of the nare group clusters was more heterogeneous and clearly different from that of the gut group. Fig. 2F shows the 10 representative microbes that contributed to the differences between the gut and the nare. These results showed a distinct microbial community composition between the gut and the nare.

3.2.2. Discovering the microbial composition differences across various physiological conditions

The microbial composition differs across diverse biological conditions [72,73]. To evaluate the capacity of CDEMI to characterize the differences in microbial composition across physiological conditions, the skin microbiome data from Mande et al. [39] (listed in Table 1)

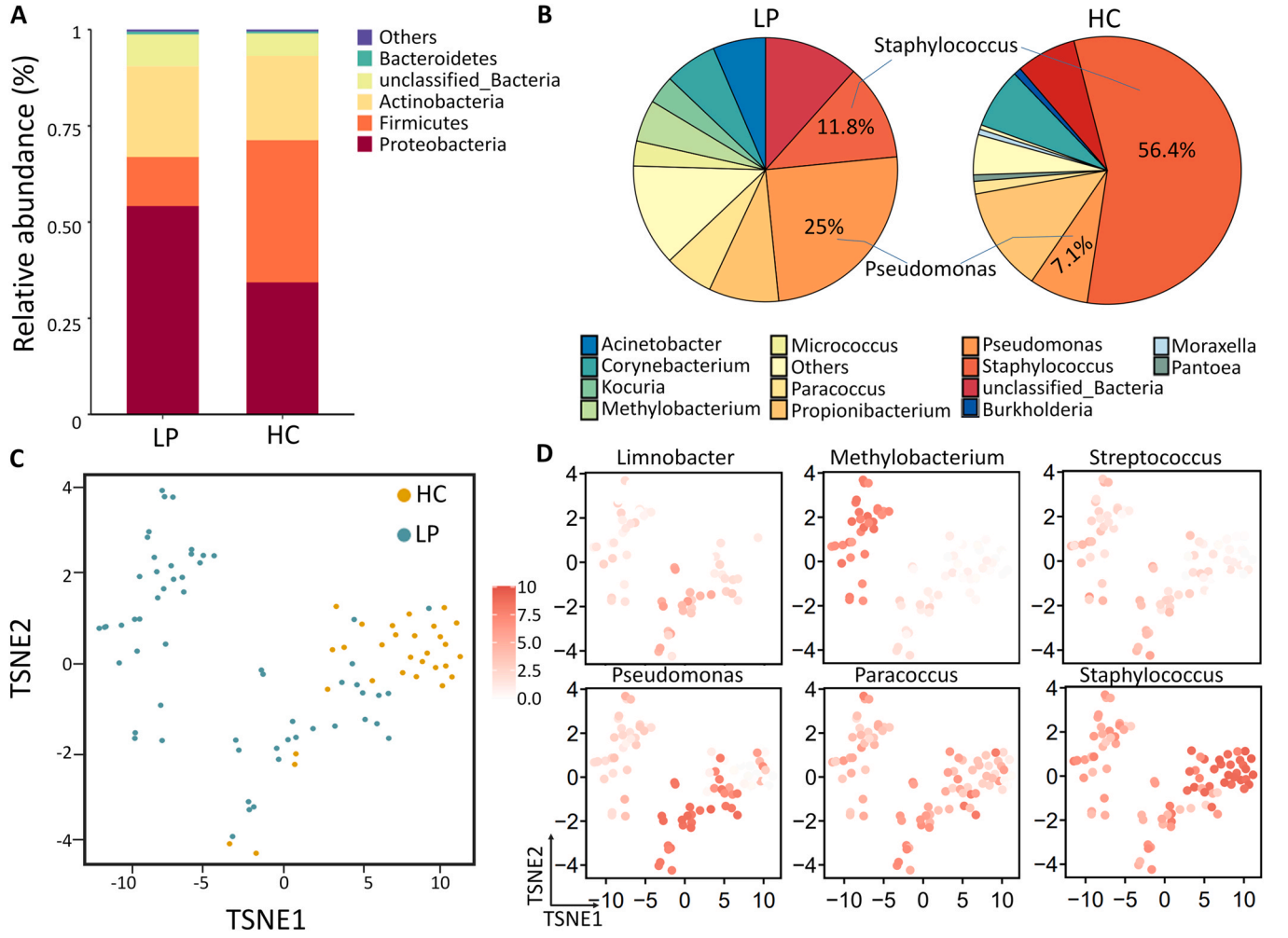


Fig. 3. Differences in microbial community composition between the LP and HC. Sample data was from Mande et al. [39]. A. The relative abundance of the 5 dominant phyla and others in the LP and HC microbiome. The less abundant phyla were grouped under "others". B. The relative abundance of the predominant genera (top ten genera) and others in the LP or HC. The less abundant genera were grouped under "others". The relative abundance of dominant microbes was calculated using the mean relative abundance for each microbe of the LP and HC groups. C. The t-SNE plot of the LP and HC microbiome. The colors represent physiological conditions (yellow: HC; blue: LP). D. The abundance distribution of representative microbes across all samples. The color key from light to dark indicates abundance levels from low to high. LP: leprosy patients; HC: healthy controls.

were applied in this case study, which contained 88 skin swab samples (58 leprosy patients and 30 healthy samples) from Hyderabad and Miraj cities in India. As shown in Fig. 3, the microbial community structure of the human skin was different between leprosy patients (LP) and healthy control (HC) groups. Fig. 3A illustrates the relative abundance of dominant phyla in the skin microbiome of the two groups using a stack column plot. The predominant phyla were largely consistent between the LP and HC groups, but their relative abundances differed substantially. Notably, *Proteobacteria* were enriched in LP samples, whereas *Firmicutes* were enriched in HC. At the genus level, each group of samples showed obvious individual differences: *Pseudomonas* accounted for 25–7.1% of the two groups, and the proportion of *Staphylococcus* in the two groups ranged from 11.8% to 56.4% (Fig. 3B).

To quantify the variation in microbial community composition under different physiological conditions, t-SNE clustering was used for 88 skin swab samples. As shown in Fig. 3C, the microbial composition showed clear differences between the LP and HC groups at the overall microbiome level. The abundance distribution of six representative microbes (discrepant between the LP and HC groups) across all microbial samples is shown in Fig. 3D. *Limnobacter* [74], *Methylobacterium* [75], *Streptococcus* [76], and *Pseudomonas* [77] were specifically enriched in the LP group, especially for *Methylobacterium*, and depleted in the HC group. These four genera were previously reported to be abnormally elevated in the LP group and are associated with human skin infections [39]. *Staphylococcus* was specifically enriched in the HC group and depleted in the LC group. These results indicate that skin microbiome composition may be associated with the physiological conditions of the host.

3.2.3. Discovering the microbial composition differences across populations within various cities

Increasing evidence shows that microbial communities exhibit distinct geography trends [39]. To evaluate the capacity of CDEMI to characterize the differences in microbial composition across populations within various cities, the skin microbiome data from Mande et al. was analyzed again. As shown in Fig. 4, the microbial community structure of human skin was different between Hyderabad and Miraj cities in India. As illustrated in Fig. 4A, the predominant phyla were largely consistent among the four groups, but their relative abundances differed substantially. The relative abundance of phyla *Proteobacteria* and *Firmicutes* in the skin microbes of the HC group exhibited clear differences between Miraj and Hyderabad but were not distinguished in the LP group of these two regions. Moreover, at a lower taxonomic level, the genus *Staphylococcus* constituted the dominant HC core taxa in the skin microbes of Hyderabad and Miraj cities, whereas the relative abundance of the genera *Pseudomonas*, *Corynebacterium*, and *Methylobacterium* in the

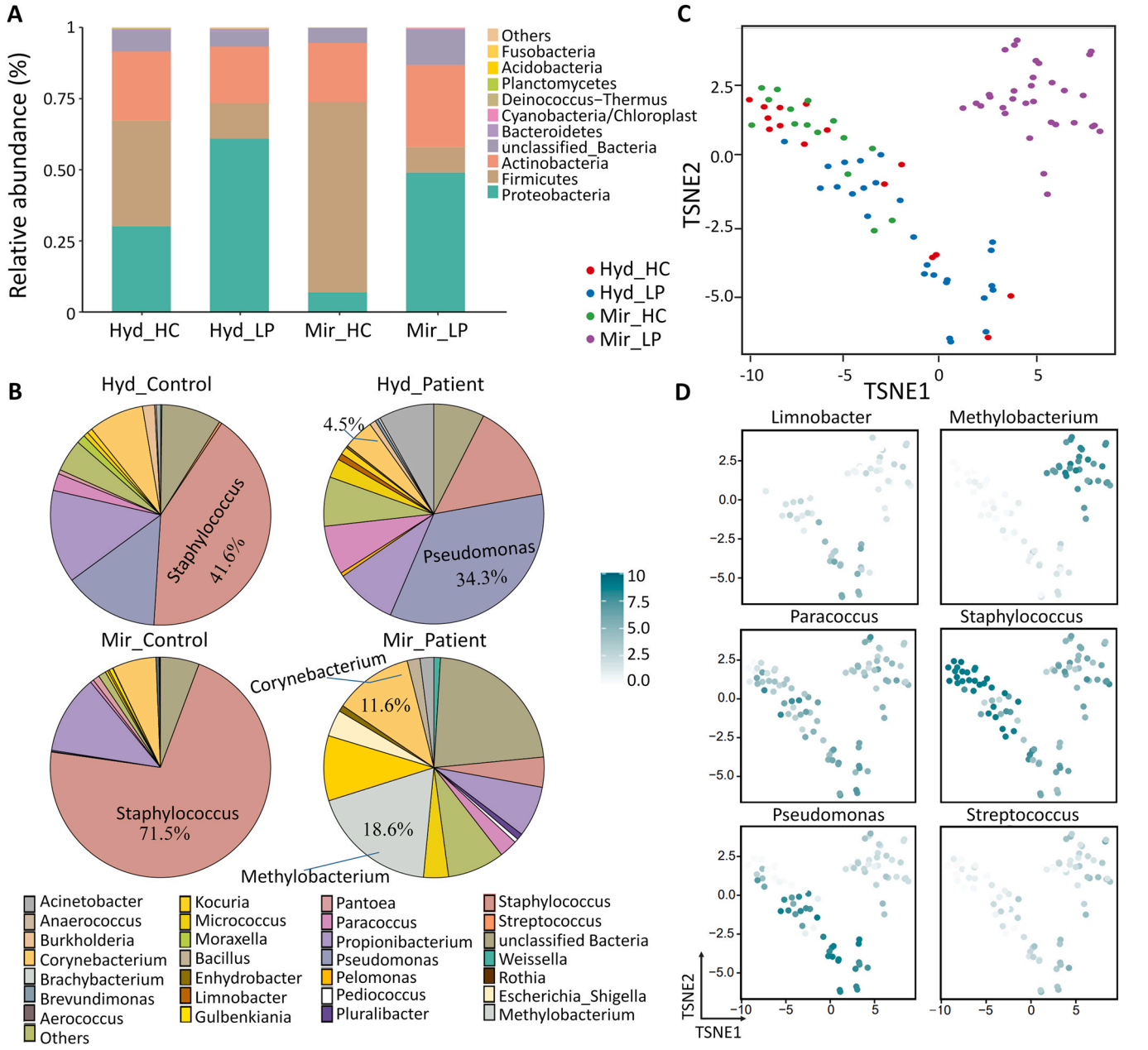


Fig. 4. The differences in microbial community composition between LP and HC groups at different geographical sampling locations. Sample data was from Mande et al. [39]. A. The relative abundance of the 10 dominant phyla and others in the LP and HC. The less abundant phyla were grouped under "others". B. The relative abundance of the predominant genera (top fifteen genera) and others in the Hyd_HC, Mir_HC, Hyd_LP, and Mir_LP groups. The less abundant genera were grouped under "others". The relative abundance of dominant microbes was calculated using the mean relative abundance for each microbe of the LP and HC groups. **C**. The t-SNE plot of the skin swab microbiome from Hyd_HC, Mir_HC, Hyd_LP, and Mir_LP groups. The colors represent geographical sampling locations under LP and HC conditions (red: Hyd_HC; blue: Hyd_LP; green: Mir_HC; purple: Mir_LP). **D**. The abundance distribution of representative microbes across all samples. The color key from light to dark indicates abundance levels from low to high. Hyd: Hyderabad; Mir: Miraj, LP: leprosy patients; HC: healthy controls.

skin microbiome of the LP group exhibited an obvious difference between Hyderabad and Miraj (Fig. 4B).

In addition, t-SNE was used to visualize the variation in microbial community composition between different cities. Fig. 4C shows that the overall skin microbiome of the Mir_LP and Hyd_LP groups exhibited distinct clusters, whereas the skin microbiomes of the Mir_HC and Hyd_HC groups could not be well separated. The skin microbiome of the Mir_LP group was different from that of the other three types. The abundance distribution of six representative skin infection-related microbes across all samples was displayed in Fig. 4D. *Methylobacterium* [75] was enriched in the Mir_LP group, but low in the Hyd_LP group, whereas *Pseudomonas* [78] was low in the Mir_LP group but enriched in the Hyd_LP group. These results suggest that the composition of the skin microbiome may be associated with the geographical location of the host.

Collectively, the above three case studies demonstrate the capability of CDEMI to characterize variations in microbial composition across various conditions (for example, geographical locations, physiological conditions, and body sites) in microbiome studies.

3.2.4. Microbial functions annotation and enrichment analysis using CDEMI

The human microbiome harbors hundreds of pathways, many of which likely modulated host biology [4], and could be an effective therapeutic target for metabolic diseases [79]. To evaluate the capacity of CDEMI in microbial function annotation and enrichment

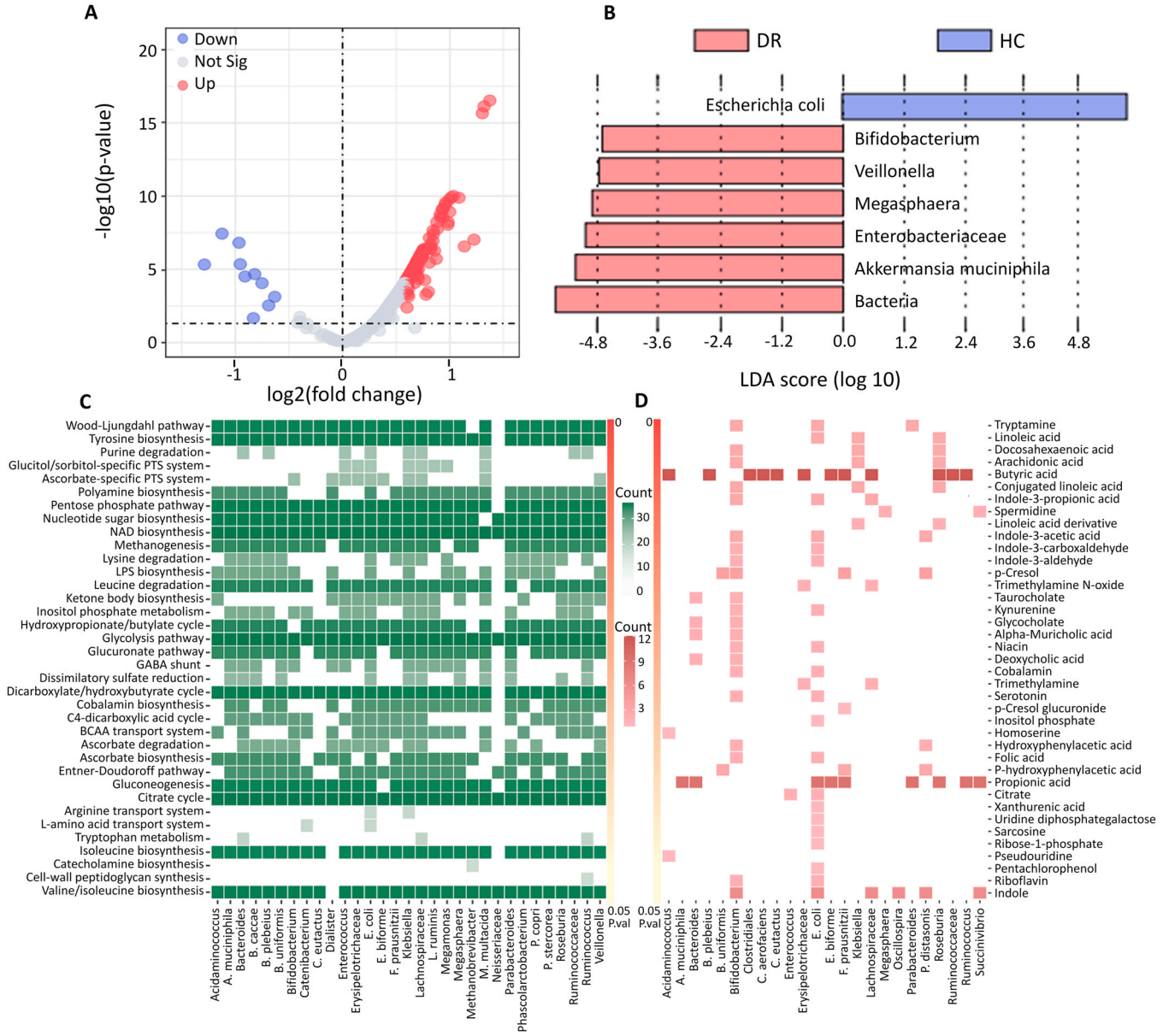


Fig. 5. Microbial annotation and enrichment analysis based on differentially abundant OTUs between the DR and HC. Sample data was from Shivaji et al. [40]. A. Volcano plot of differentially abundant OTUs between the DR and HC (|logFC| > 0.585, *p*-value < 0.05), eOTUs: enriched OTUs, dOTUs: depleted OTUs, oOTUs: ordinary OTUs. B. LDA scores of the differentially abundant OTUs between the DR and HC. LDA scores were generated from the LEfSe analysis (LDA > 2.0, *p*-value < 0.05). One bacterial OTU was enriched in HC and 6 OTUs were enriched in DR. C. Functional enrichment analysis results were based on the microbial function library in the CDEMI. Colors represent the counts of microbes involved in this pathway. D. Microbe-derived metabolite enrichment analysis results. Colors represent the counts of microbes associated with this metabolite. BCAA transport system: Branched-chain amino acid transport system; Hydroxypropionate/butylate cycle: Hydroxypropionate-hydroxybutylate cycle; GABA shunt: GABA (gamma-Aminobutyrate) shunt; LPS: Lipopolysaccharide; Glucitol/sorbitol-specific PTS system: PTS system, glucitol/sorbitol-specific II component; Ascorbate-specific PTS system: PTS system, ascorbate-specific II component; DR: diabetic retinopathy; HC: healthy controls.

analysis, fecal microbiome data from Shivaji et al.[40] were used, which contained 3539 Operational Taxonomic Units (OTUs) from 28 diabetic retinopathy (DR) samples and 30 healthy controls (HC). In particular, differentially abundant OTUs between the DR and HC groups were identified using edgeR packages [80]. Enriched OTUs (eOTUs) and depleted OTUs (dOTUs) specifically represented OTUs that were more than 0.585 times higher or lower in relative abundance (P < 0.05) in the DR samples than in the HC samples. As illustrated in Fig. 5A, 143 OTUs were identified as significantly different between the DR and HC groups (Supplementary Table S1), which included 133 eOTUs and 10 dOTUs. The top discriminatory OTUs are listed by the linear discriminant analysis Effect Size tool (LEfSe) [81] in Fig. 5B. *Bifidobacterium* was more abundant in DR samples, which is consistent with the results of Ma et al. [82]. Moreover, based on the metabolic pathways from the microbial function library, we performed metabolic pathway enrichment analysis for the differential OTUs between the DR and HC groups. The significantly enriched pathways (for example, glycolysis pathway [83], valine/isoleucine biosynthesis [84], cell-wall peptidoglycan synthesis [85], and tryptophan metabolism [86]) are listed in Supplementary Table S2, and descriptions of the relevance between DR and pathways are shown in Supplementary Table S3. The microbes involved in these pathways are shown in Fig. 5C.

3.2.5. Characterizing metabolites produced/synthesized by microbiota using CDEMI

Diverse microbial communities modulate host phenotypes through the production of small molecules (for example, bile acids,

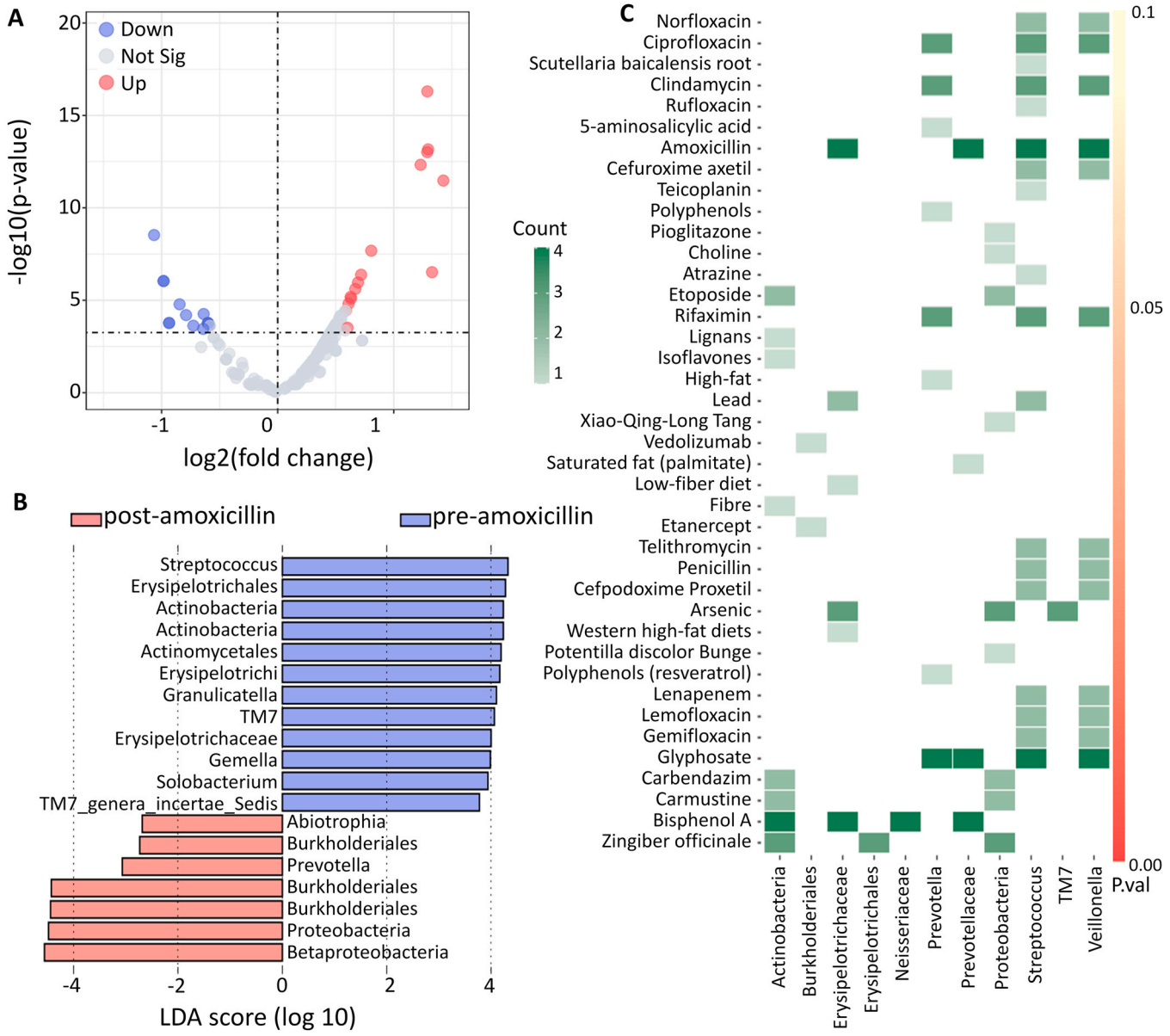


Fig. 6. Microbial annotation and enrichment analysis based on differentially abundant OTUs between the pre-amoxicillin and post-amoxicillin groups. Sample data from Schrenzel et al. [41]. A. Volcano plot of differentially abundant OTUs between the pre-amoxicillin and post-amoxicillin groups (|logFC| > 0.585 and p-value < 0.05), eOTUs: enriched OTUs, dOTUs: depleted OTUs, oOTUs: ordinary OTUs. B. LDA scores of the differentially abundant OTUs between the pre-amoxicillin and post-amoxicillin groups. LDA scores were generated from the LEfSe analysis (LDA > 2.0, p-value < 0.05). 12 OTUs were enriched in the pre-amoxicillin group and 7 were enriched in the post-amoxicillin group. C. EASs enrichment analysis results based on microbe-associated EASs library in CDEMI. Colors represented the counts of microbes involved in EASs.

short-chain fatty acids, and indole derivatives) [87] and have been considered to be related to the pathogenesis of metabolic disorders [88]. To evaluate the capacity of CDEMI to investigate how microbes modulate host phenotypes through the production of small molecules, fecal microbiome data from Shivaji et al. were used. Based on the bioactive microbial metabolite library in CDEMI, we performed a microbe-derived metabolite enrichment analysis for the 143 differential OTUs between the DR and HC groups. The DR-associated metabolites (for example, indole, arachidonic acid, and trimethylamine N-oxide) [86,89,90] are listed in Supplementary Table S4, and descriptions of the relevance between DR and metabolite are shown in Supplementary Table S5. The associations between the microbes and the derived metabolites are shown in Fig. 5D. In summary, these results demonstrate the capability of CDEMI for microbial function annotation and characterization of MDMs.

3.2.6. Characterizing the potential EAS interacting with microbe by CDEMI

The interaction between EAS and microbes plays a crucial role in human health, disease, and physiological responses to diverse clues and treatments [12]. To evaluate the capacity of CDEMI to characterize the potential exogenous active substances interacting with microbes, the salivary microbiome data from Schrenzel et al. [41] was applied, which contained 1656 OTUs from 36 acute otitis media samples before and after amoxicillin treatment. In particular, the differential abundance of microbes between the pre-amoxicillin and post-amoxicillin groups was identified by edgeR [80]. As shown in Fig. 6A, 32 significantly different OTUs with a p-value < 0.05 were identified between the pre-amoxicillin and post-amoxicillin groups (Supplementary Table S6), including 18 enriched OTUs (eOTUs) and 14 depleted OTUs (dOTUs). The top discriminatory OTUs are listed

using LEfSe [74] in Fig. 6B. Amoxicillin treatment resulted in a reduction in the abundance of the phyla *TM7* and *Actinobacteria*, and the genus *Streptococcus*, which was consistent with the findings of Schrenzel et al.[41]. Moreover, based on the EASs library in CDEMI, we performed EASs enrichment analysis on the differential OTUs between the pre-amoxicillin and post-amoxicillin groups and found that most EASs are involved in various antibiotics that possess anti-inflammatory potencies (for example, gemifloxacin [91], penicillins [92], and lemofloxacin [93]) (Supplementary Table S7), and inflammation associated active substances (for example, arsenic, glyphosate, and bisphenol A) [94–96]. The relationship between EASs and inflammation is shown in Supplementary Table S8. The associations between the microbes and EASs are shown in Fig. 6C. In conclusion, these results suggest that EASs has a potential impact on the microbial community composition.

4. Conclusions

Collectively, CDEMI is distinguished for its capacity to characterize the differences in microbial community composition and function from five microbe libraries integrated for microbiome study, and it is expected to emerge as an indispensable complement to other available tools. With the emergence of large metagenomic and 16 S rRNA sequencing data, CDEMI could be used to investigate how exogenous substances (for example, nutritional and environmental substances) affect microbial community distribution, composition, and function, discover the related mechanisms involved in mediating human diseases, and facilitate the development of a treatment strategy that improves human health.

CRediT authorship contribution statement

Lidan Wang: Visualization, Data curation, Software, Writing – original draft, Writing – review & editing. **Xiao Liang**: Visualization, Data curation, Software. **Hao Chen**: Data curation. **Lijie Cao**: Data curation. **Lan Liu**: Data curation. **Feng Zhu**: Conceptualization, Project administration. **Yubin Ding**: Supervision, Funding acquisition. **Jing Tang**: Funding acquisition, Supervision, Writing – original draft, Writing – review & editing, Conceptualization. **Youlong Xie**: Conceptualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2018YFC1004401), the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202100421), the Basic Research and Frontiers Exploration Project of the Science and Technology Committee of Yuzhong District, Chongqing (20210119), and the Chongqing Science and Technology Commission (cstc2021jcyj-bshX0105).

Author contributions

J.T. conceived the idea and supervised the work. L.W. and Y.X. performed the research, and data curation. L.W., X.L., and Y.X. implemented the visualization, constructed the web server, and wrote the scripts. X.L., H.C., L.C., L.L., F.Z., and Y.D. prepared and analyzed the data. J.T. and L.W. wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.03.044.

References

[1] Ranallo RT, McDonald LC, Halpin AL, Hiltke T, Young VB. The state of microbiome science at the intersection of infectious diseases and antimicrobial resistance. J Infect Dis 2021;223:S187–93. https://doi.org/10.1093/infdis/jiab020
[2] Han S, Van Treuren W, Fischer CR, Merrill BD, DeFelice BC, et al. A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. Nature 2021;595:415–20. https://doi.org/10.1038/s41586-021-03707-9
[3] Shine EE, Crawford JM. Molecules from the microbiome. Annu Rev Biochem 2021;90:789–815. https://doi.org/10.1146/annurev-biochem-080320-115307
[4] Funabashi M, Grove TL, Wang M, Varma Y, McFadden ME, et al. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. Nature 2020;582:566–70. https://doi.org/10.1038/s41586-020-2396-4
[5] Krautkramer KA, Fan J, Backhed F. Gut microbial metabolites as multi-kingdom intermediates. Nat Rev Microbiol 2021;19:77–94. https://doi.org/10.1038/s41579-020-0438-4
[6] Martins MD, Jiao Y, Larsson L, Almeida LO, Garaicoa-Pazmino C, et al. Epigenetic modifications of histones in periodontal disease. J Dent Res 2016;95:215–22. https://doi.org/10.1177/0022034515611876
[7] Bierne H, Pourpre R. Bacterial factors targeting the nucleus: the growing family of nucleomodulins. Toxins 2020;12:220. https://doi.org/10.3390/toxins12040220
[8] Song M, Chan AT, Sun J. Influence of the gut microbiome, diet, and environment on risk of colorectal cancer. Gastroenterology 2020;158:322–40. https://doi.org/10.1053/j.gastro.2019.06.048
[9] Weersma RK, Zhernakova A, Fu J. Interaction between drugs and the gut microbiome. Gut 2020;69:1510–9. https://doi.org/10.1136/gutjnl-2019-320204
[10] Zhang X, Yang Y, Zhang F, Yu J, Sun W, et al. Traditional Chinese medicines differentially modulate the gut microbiota based on their nature (Yao-Xing). Phytomedicine 2021;85:153496. https://doi.org/10.1016/j.phymed.2021.153496
[11] Ruuskanen S, Rainio MJ, Gomez-Gallego C, Selenius O, Salminen S, et al. Glyphosate-based herbicides influence antioxidants, reproductive hormones and gut microbiome but not reproduction: a long-term experiment in an avian model. Environ Pollut 2020;266:115108. https://doi.org/10.1016/j.envpol.2020.115108
[12] Zeng X, Yang X, Fan J, Tan Y, Ju L, et al. MASI: microbiota-active substance interactions database. Nucleic Acids Res 2021;49:D776–82. https://doi.org/10.1093/nar/gkaa924
[13] Sun H, Huang X, Huo B, Tan Y, He T, et al. Detecting sparse microbial association signals adaptively from longitudinal microbiome data based on generalized estimating equations. Brief Bioinform 2022;23:bbac149. https://doi.org/10.1093/bib/bbac149
[14] Grice EA, Segre JA. The skin microbiome. Nat Rev Microbiol 2011;9:244–53. https://doi.org/10.1038/nrmicro2537
[15] DeJong EN, Surette MG, Bowdish DME. The gut microbiota and unhealthy aging: disentangling cause from consequence. Cell Host Microbe 2020;28:180–9. https://doi.org/10.1016/j.chom.2020.07.013
[16] Caparros E, Wiest R, Scharl M, Rogler G, Gutierrez Casbas A, et al. Dysbiotic microbiota interactions in Crohn's disease. Gut Microbes 2021;13:1949096. https://doi.org/10.1080/19490976.2021.1949096
[17] Tropini C, Earle KA, Huang KC, Sonnenburg JL. The gut microbiome: connecting spatial organization to function. Cell Host Microbe 2017;21:433–42. https://doi.org/10.1016/j.chom.2017.03.010
[18] Neugent ML, Hulyalkar NV, Nguyen VH, Zimmern PE, De Nisco NJ. Advances in understanding the human urinary microbiome and its potential role in urinary tract infection. mBio 2020;11:e00218–20. https://doi.org/10.1128/mBio.00218-20
[19] Derrien M, Alvarez AS, de Vos WM. The gut microbiota in the first decade of life. Trends Microbiol 2019;27:997–1010. https://doi.org/10.1016/j.tim.2019.08.001
[20] Munoz A, Hayward MR, Bloom SM, Rocafort M, Ngcapu S, et al. Correction to: modeling the temporal dynamics of cervicovaginal microbiota identifies targets that may promote reproductive health. Microbiome 2021;9:206. https://doi.org/10.1186/s40168-021-01171-1
[21] Lynch JB, Hsiao EY. Microbiomes as sources of emergent host phenotypes. Science 2019;365:1405–9. https://doi.org/10.1126/science.aay0240
[22] Zhang Q, Yu K, Li S, Zhang X, Zhao Q, et al. gutMEGA: a database of the human gut MEtaGenome Atlas. Brief Bioinform 2021;22:bbaa082. https://doi.org/10.1093/bib/bbaa082
[23] Liu ZQ, Ma AJ, Mathe E, Merling M, Ma Q, et al. Network analyses in microbiome based on high-throughput multi-omics data. Brief Bioinform 2021;22:1639–55. https://doi.org/10.1093/bib/bbaa005
[24] Yang Y. Emerging patterns of microbial functional traits. Trends Microbiol 2021;29:874–82. https://doi.org/10.1016/j.tim.2021.04.004
[25] Zou A, Nadeau K, Xiong X, Wang PW, Copeland JK, et al. Systematic profiling of the chicken gut microbiome reveals dietary supplementation with antibiotics alters expression of multiple microbial pathways with minimal impact on community structure. Microbiome 2022;10:127. https://doi.org/10.1186/s40168-022-01319-7

[26] Ma W, Huang C, Zhou Y, Li J, Cui Q. MicroPattern: a web-based tool for microbe set enrichment analysis and disease similarity calculation based on a list of microbes. Sci Rep 2017;7:40200. https://doi.org/10.1038/srep40200
[27] Kou Y, Xu X, Zhu Z, Dai L, Tan Y. Microbe-set enrichment analysis facilitates functional interpretation of microbiome profiling data. Sci Rep 2020;10:21466. https://doi.org/10.1038/s41598-020-78511-y
[28] Loos D, Zhang L, Beemelmanns C, Kurzai O, Panagiotou G. DAnIEL: a user-friendly web server for fungal ITS amplicon sequencing data. Front Microbiol 2021;12:720513. https://doi.org/10.3389/fmicb.2021.720513
[29] Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017;45:D353–61. https://doi.org/10.1093/nar/gkw1092
[30] Le Boulch M, Dehais P, Combes S, Pascal G. The MACADAM database: a MetAboliC pAthways DAtabase for Microbial taxonomic groups for mining potential metabolic capacities of archaeal and bacterial taxonomic groups. Database 2019;2019:baz049. https://doi.org/10.1093/database/baz049
[31] Forster SC, Browne HP, Kumar N, Hunt M, Denise H, et al. HPMCD: the database of human microbial communities from metagenomic datasets and microbial reference genomes. Nucleic Acids Res 2016;44:D604–9. https://doi.org/10.1093/nar/gkv1216
[32] Zhu C, Mahlich Y, Miller M, Bromberg Y. fusionDB: assessing microbial diversity and environmental preferences via functional similarity networks. Nucleic Acids Res 2018;46:D535–41. https://doi.org/10.1093/nar/gkx1060
[33] Dahan E, Martin VM, Yassour M. EasyMap - an interactive web tool for evaluating and comparing associations of clinical variables and microbiome composition. Front Cell Infect Microbiol 2022;12:854164. https://doi.org/10.3389/fcimb.2022.854164
[34] Mori H, Maruyama F, Kurokawa K. VITCOMIC: visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences. BMC Bioinform 2010;11:332. https://doi.org/10.1186/1471-2105-11-332
[35] Liu Z, Mi K, Xu ZZ, Zhang Q, Liu X. PM2RA: a framework for detecting and quantifying relationship alterations in microbial community. Genom Proteom Bioinform 2021;19:154–67. https://doi.org/10.1016/j.gpb.2020.07.005
[36] Dhariwal A, Chong J, Habib S, King IL, Agellon LB, et al. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res 2017;45:W180–8. https://doi.org/10.1093/nar/gkx295
[37] Anand S, Kuntal BK, Mohapatra A, Bhatt V, Mande SS. FunGeCo: a web-based tool for estimation of functional potential of bacterial genomes and microbiomes using gene context information. Bioinformatics 2020;36:2575–7. https://doi.org/10.1093/bioinformatics/btz957
[38] Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, et al. Longitudinal multi-omics of host-microbe dynamics in prediabetes. Nature 2019;569:663–71. https://doi.org/10.1038/s41586-019-1236-x
[39] Bayal N, Nagpal S, Haque MM, Patole MS, Shouche Y, et al. Structural aspects of lesional and non-lesional skin microbiota reveal key community changes in leprosy patients from India. Sci Rep 2021;11:3294. https://doi.org/10.1038/s41598-020-80533-5
[40] Das T, Jayasudha R, Chakravarthy S, Prashanthi GS, Bhargava A, et al. Alterations in the gut bacterial microbiome in people with type 2 diabetes mellitus and diabetic retinopathy. Sci Rep 2021;11:2738. https://doi.org/10.1038/s41598-021-82538-0
[41] Lazarevic V, Manzano S, Gaia N, Girard M, Whiteson K, et al. Effects of amoxicillin treatment on the salivary microbiota in children with acute otitis media. Clin Microbiol Infect 2013;19:E335–42. https://doi.org/10.1111/1469-0691.12213
[42] Sarkar A, Harty S, Johnson KV, Moeller AH, Carmody RN, et al. The role of the microbiome in the neurobiology of social behaviour. Biol Rev Camb Philos Soc 2020;95:1131–66. https://doi.org/10.1111/brv.12603
[43] Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, et al. Application of a neutral community model to assess structuring of the human lung microbiome. e02284-14 mBio 2015;6. https://doi.org/10.1128/mBio.02284-14
[44] Leech T, McDowall L, Hopkins KP, Sait SM, Harrison XA, et al. Social environment drives sex and age-specific variation in Drosophila melanogaster microbiome composition and predicted function. Mol Ecol 2021;30:5831–43. https://doi.org/10.1111/mec.16149
[45] van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579–605.
[46] Perez Rosero E, Heron S, Jovel J, O'Neil CR, Turvey SL, et al. Differential signature of the microbiome and neutrophils in the oral cavity of HIV-infected individuals. Front Immunol 2021;12:780910. https://doi.org/10.3389/fimmu.2021.780910
[47] Coe RKR. Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies.
[48] Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, et al. The virtual metabolic human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Res 2019;47:D614–24. https://doi.org/10.1093/nar/gky992
[49] Wang L, Zhang W, Wu X, Liang X, Cao L, et al. MIAOME: human microbiome affect the host epigenome. Comput Struct Biotechnol J 2022;20:2455–63. https://doi.org/10.1016/j.csbj.2022.05.024
[50] Tang J, Wu X, Mou M, Wang C, Wang L, et al. GIMICA: host genetic and immune factors shaping human microbiota. Nucleic Acids Res 2021;49:D715–22. https://doi.org/10.1093/nar/gkaa851
[51] Cheng L, Qi C, Zhuang H, Fu T, Zhang X. gutMDisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. Nucleic Acids Res 2020;48:D554–60. https://doi.org/10.1093/nar/gkz843
[52] Janssens Y, Nielandt J, Bronselaer A, Debunne N, Verbeke F, et al. Disbiome database: linking the microbiome to disease. BMC Microbiol 2018;18:50. https://doi.org/10.1186/s12866-018-1197-5
[53] Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2021;49:D10–7. https://doi.org/10.1093/nar/gkaa892
[54] Meng X, Zhou HY, Shen HH, Lufumpa E, Li XM, et al. Microbe-metabolite-host axis, two-way action in the pathogenesis and treatment of human autoimmunity. Autoimmun Rev 2019;18:455–75. https://doi.org/10.1016/j.autrev.2019.03.006
[55] Van Treuren W, Dodd D. Microbial contribution to the human metabolome: implications for health and disease. Annu Rev Pathol 2020;15:345–69. https://doi.org/10.1146/annurev-pathol-020117-043559
[56] Kim S, Chen J, Cheng T, Gindulyte A, He J, et al. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res 2021;49:D1388–95. https://doi.org/10.1093/nar/gkaa971
[57] Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 2018;46:D608–17. https://doi.org/10.1093/nar/gkx1089
[58] Kazemian N, Mahmoudi M, Halperin F, Wu JC, Pakpour S. Gut microbiota and cardiovascular disease: opportunities and challenges. Microbiome 2020;8:36. https://doi.org/10.1186/s40168-020-00821-0
[59] Stewart OA, Wu F, Chen Y. The role of gastric microbiota in gastric cancer. Gut Microbes 2020;11:1220–30. https://doi.org/10.1080/19490976.2020.1762520
[60] Miyauchi E, Shimokawa C, Steimle A, Desai MS, Ohno H. The impact of the gut microbiome on extra-intestinal autoimmune diseases. Nat Rev Immunol 2023;23:9–23. https://doi.org/10.1038/s41577-022-00727-y
[61] Sekirov I, Russell SL, Antunes LC, Finlay BB. Gut microbiota in health and disease. Physiol Rev 2010;90:859–904. https://doi.org/10.1152/physrev.00045.2009
[62] Montassier E, Al-Ghalith GA, Hillmann B, Viskocil K, Kabage AJ, et al. CLOUD: a non-parametric detection test for microbiome outliers. Microbiome 2018;6:137. https://doi.org/10.1186/s40168-018-0514-4
[63] Guan TP, Teng JLL, Fong JYH, Lau SKP, Woo PCY. Seasonal shift in gut microbiome diversity in wild Sichuan takin (Budorcas tibetanus) and environmental adaptation. Comput Struct Biotechnol J 2023;21:1283–91. https://doi.org/10.1016/j.csbj.2022.12.035
[64] Doolittle JM, Webster-Cyriaque J. Polymicrobial infection and bacterium-mediated epigenetic modification of DNA tumor viruses contribute to pathogenesis. e01015-01014 mBio 2014;5. https://doi.org/10.1128/mBio.01015-14
[65] Gao L, Xu T, Huang G, Jiang S, Gu Y, et al. Oral microbiomes: more and more importance in oral cavity and whole body. Protein Cell 2018;9:488–500. https://doi.org/10.1007/s13238-018-0548-1
[66] Trischler R, Roth J, Sorbara MT, Schlegel X, Muller V. A functional Wood-Ljungdahl pathway devoid of a formate dehydrogenase in the gut acetogens Blautia wexlerae, Blautia luti and beyond. Environ Microbiol 2022;24:3111–23. https://doi.org/10.1111/1462-2920.16029
[67] Verhoog S, Taneri PE, Roa Diaz ZM, Marques-Vidal P, Troup JP, et al. Dietary factors and modulation of bacteria strains of Akkermansia muciniphila and Faecalibacterium prausnitzii: a systematic Review. Nutrients 2019;11:1565. https://doi.org/10.3390/nu11071565
[68] Zafar H, Saier Jr. MH. Gut bacteroides species in health and disease. Gut Microbes 2021;13:1–20. https://doi.org/10.1080/19490976.2020.1848158
[69] Camarinha-Silva A, Jauregui R, Pieper DH, Wos-Oxley ML. The temporal dynamics of bacterial communities across human anterior nares. Environ Microbiol Rep 2012;4:126–32. https://doi.org/10.1111/j.1758-2229.2011.00313.x
[70] Yan M, Pamp SJ, Fukuyama J, Hwang PH, Cho DY, et al. Nasal microenvironments and interspecific interactions influence nasal microbiota complexity and S. aureus carriage. Cell Host Microbe 2013;14:631–40. https://doi.org/10.1016/j.chom.2013.11.005
[71] Kwiecinski JM, Horswill AR. Staphylococcus aureus bloodstream infections: pathogenesis and regulatory mechanisms. Curr Opin Microbiol 2020;53:51–60. https://doi.org/10.1016/j.mib.2020.02.005
[72] Pittayanon R, Lau JT, Leontiadis GI, Tse F, Yuan Y, et al. Differences in gut microbiota in patients with vs without inflammatory bowel diseases: a systematic review. Gastroenterology 2020;158(930–946):e1. https://doi.org/10.1053/j.gastro.2019.11.294
[73] Qi Y, Wan Y, Li T, Zhang M, Song Y, et al. Comparison of the ocular microbiomes of dry eye patients with and without autoimmune disease. Front Cell Infect Microbiol 2021;11:716867. https://doi.org/10.3389/fcimb.2021.716867
[74] Zheng Y, Wang Q, Ma L, Chen Y, Gao Y, et al. Alterations in the skin microbiome are associated with disease severity and treatment in the perioral zone of the skin of infants with atopic dermatitis. Eur J Clin Microbiol Infect Dis 2019;38:1677–85. https://doi.org/10.1007/s10096-019-03598-9
[75] Bayal N, Nagpal S, Haque MM, Patole MS, Valluri V, et al. 16S rDNA based skin microbiome data of healthy individuals and leprosy patients from India. Sci Data 2019;6:225. https://doi.org/10.1038/s41597-019-0232-1
[76] Hirose Y, Yamaguchi M, Sumitomo T, Nakata M, Hanada T, et al. Streptococcus pyogenes upregulates arginine catabolism to exert its pathogenesis on the skin surface. Cell Rep 2021;34:108924. https://doi.org/10.1016/j.celrep.2021.108924
[77] Kim TH, Li XH, Lee JH. Alleviation of Pseudomonas aeruginosa infection by propeptide-mediated inhibition of protease IV. Microbiol Spectr 2021;9:e0078221. https://doi.org/10.1128/Spectrum.00782-21
[78] Silva PE, Costa PS, Avila MP, Suhadolnik ML, Reis MP, et al. Leprous lesion presents enrichment of opportunistic pathogenic bacteria. SpringerPlus 2015;4:187. https://doi.org/10.1186/s40064-015-0955-1

[79] Wu J, Wang K, Wang X, Pang Y, Jiang C. The role of the gut microbiome and its metabolites in metabolic diseases. Protein Cell 2021;12:360–73. https://doi.org/10.1007/s13238-020-00814-7
[80] Calgaro M, Romualdi C, Waldron L, Risso D, Vitulo N. Assessment of statistical methods from single cell, bulk RNA-seq, and metagenomics applied to microbiome data. Genome Biol 2020;21:191. https://doi.org/10.1186/s13059-020-02104-1
[81] Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 2018;46:W537–44. https://doi.org/10.1093/nar/gky379
[82] Huang Y, Wang Z, Ma H, Ji S, Chen Z, et al. Dysbiosis and implication of the gut microbiota in diabetic retinopathy. Front Cell Infect Microbiol 2021;11:646348. https://doi.org/10.3389/fcimb.2021.646348
[83] Yumnamcha T, Guerra M, Singh LP, Ibrahim AS. Metabolic dysregulation and neurovascular dysfunction in diabetic retinopathy. Antioxidants 2020;9:1244. https://doi.org/10.3390/antiox9121244
[84] Wang H, Fang J, Chen F, Sun Q, Xu X, et al. Metabolomic profile of diabetic retinopathy: a GC-TOFMS-based approach using vitreous and aqueous humor. Acta Diabetol 2020;57:41–51. https://doi.org/10.1007/s00592-019-01363-0
[85] Floyd JL, Grant MB. The gut-eye axis: lessons learned from murine models. Ophthalmol Ther 2020;9:499–513. https://doi.org/10.1007/s40123-020-00278-2
[86] Guo C, Jiang D, Xu Y, Peng F, Zhao S, et al. High-coverage serum metabolomics reveals metabolic pathway dysregulation in diabetic retinopathy: a propensity score-matched study. Front Mol Biosci 2022;9:822647. https://doi.org/10.3389/fmolb.2022.822647
[87] Krautkramer KA, Kreznar JH, Romano KA, Vivas EI, Barrett-Wilt GA, et al. Diet-microbiota interactions mediate global epigenetic programming in multiple host tissues. Mol Cell 2016;64:982–92. https://doi.org/10.1016/j.molcel.2016.10.025
[88] Agus A, Clement K, Sokol H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. Gut 2021;70:1174–82. https://doi.org/10.1136/gutjnl-2020-323071
[89] Alarcon Yempen RE, Venzel R, Paulino Campos MC, de Oliveira LP, Lins RVD, et al. Gut microbiota: a potential therapeutic target for management of diabetic retinopathy. Life Sci 2021;286:120060. https://doi.org/10.1016/j.lfs.2021.120060
[90] Ye P, Zhang X, Xu Y, Xu J, Song X, et al. Alterations of the gut microbiome and metabolome in patients with proliferative diabetic retinopathy. Front Microbiol 2021;12:667632. https://doi.org/10.3389/fmicb.2021.667632
[91] AlKhalil M, Al-Hiari Y, Kasabri V, Arabiyat S, Al-Zweiri M, et al. Selected pharmacotherapy agents as antiproliferative and anti-inflammatory compounds. Drug Dev Res 2020;81:470–90. https://doi.org/10.1002/ddr.21640
[92] Wilhelmus KR, Yokoyama CM. Syphilitic episcleritis and scleritis. Am J Ophthalmol 1987;104:595–7. https://doi.org/10.1016/0002-9394(87)90170-x
[93] Zhu J, Qiu J, Chen K, Wang W, Zheng S. Tea polyphenols and Levofloxacin alleviate the lung injury of hepatopulmonary syndrome in common bile duct ligation rats through endotoxin -TNF signaling. Biomed Pharmacother 2021;137:111263. https://doi.org/10.1016/j.biopha.2021.111263
[94] Farkhondeh T, Samarghandian S, Azimi-Nezhad M. The role of arsenic in obesity and diabetes. J Cell Physiol 2019;234:12516–29. https://doi.org/10.1002/jcp.28112
[95] Loffredo LF, Coden ME, Berdnikovs S. Endocrine disruptor bisphenol A (BPA) triggers systemic para-inflammation and is sufficient to induce airway allergic sensitization in mice. Nutrients 2020;12:343. https://doi.org/10.3390/nu12020343
[96] Pandher U, Kirychuk S, Schneberger D, Thompson B, Aulakh G, et al. Pulmonary inflammatory response from co-exposure to LPS and glyphosate. Environ Toxicol Pharmacol 2021;86:103651. https://doi.org/10.1016/j.etap.2021.103651