Microbiome Search Engine 2: a Platform for Taxonomic and Functional Search of Global Microbiomes on the Whole-Microbiome Level

Gongchao Jing, Lu Liu, Zengbin Wang, Yufeng Zhang, Li Qian, Chunxia Gao, Meng Zhang, Min Li, Zhenkun Zhang, Xiaohan Liu, Jian Xu, Xiaoquan Su

aCollege of Computer Science and Technology, Qingdao University, Qingdao, China
bSingle-Cell Center, CAS Key Laboratory of Biofuels and Shandong Key Laboratory of Energy Genetics, Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China
cShandong Institute of Energy Research, Qingdao, China
dNational Science Library, Chinese Academy of Sciences, Beijing, China
eDepartment of Computer Science and Technology, Ocean University of China, Qingdao, China
fKey Laboratory of Marine Drugs, the Ministry of Education of China, School of Medicine and Pharmacy, Ocean University of China, Qingdao, China

ABSTRACT Metagenomic data sets from diverse environments have been growing rapidly. To ensure accessibility and reusability, tools that quickly and informatively correlate new microbiomes with existing ones are in demand. Here, we introduce Microbiome Search Engine 2 (MSE 2), a microbiome database platform for searching query microbiomes in the global metagenome data space based on the taxonomic or functional similarity of a whole microbiome to those in the database. MSE 2 consists of (i) a well-organized and regularly updated microbiome database that currently contains over 250,000 metagenomic shotgun and 16S rRNA gene amplicon samples associated with unified metadata collected from 798 studies, (ii) an enhanced search engine that enables real-time and fast (<0.5 s per query) searches against the entire database for best-matched microbiomes using overall taxonomic or functional profiles, and (iii) a Web-based graphical user interface for user-friendly searching, data browsing, and tutoring. MSE 2 is freely accessible via http://mse.ac.cn. For standalone searches of customized microbiome databases, the kernel of the MSE 2 search engine is provided at GitHub (https://github.com/qibebt-bioinfo/meta-storms).

IMPORTANCE A search-based strategy is useful for large-scale mining of microbiome data sets, such as a bird’s-eye view of the microbiome data space and disease diagnosis via microbiome big data. Here, we introduce Microbiome Search Engine 2 (MSE 2), a microbiome database platform for searching query microbiomes against the existing microbiome data sets on the basis of their similarity in taxonomic structure or functional profile. Key improvements include database extension, data compatibility, a search engine kernel, and a user interface. The new ability to search the microbiome space via functional similarity greatly expands the scope of search-based mining of the microbiome big data.

KEYWORDS amplicon, metagenome, microbiome, online service, search engine

Metagenomic approaches have been widely employed to probe microbiomes among various habitats by linking dynamics of microbial compositions and predicted functions to environmental changes (1, 2), human disease development (3–7), and drug responses (8, 9). With the rapid development of sampling strategies and sequencing technologies, an enormous volume of microbiome data sets, including

Citation Jing G, Liu L, Wang Z, Zhang Y, Qian L, Gao C, Zhang M, Li M, Zhang Z, Liu X, Xu J, Su X. 2021. Microbiome Search Engine 2: a platform for taxonomic and functional search of global microbiomes on the whole-microbiome level. mSystems 6:e00943-20. https://doi.org/10.1128/mSystems.00943-20.

Editor Morgan G. I. Langille, Dalhousie University
Copyright © 2021 Jing et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.
Address correspondence to Jian Xu, xujian@qibebt.ac.cn, or Xiaoquan Su, suxq@qdu.edu.cn.
Received 17 September 2020
Accepted 23 December 2020
Published 19 January 2021
both 16S rRNA gene amplicon-based and shotgun whole-genome sequencing (WGS)-
based data sets, has been produced by individual, small-cohort projects or large-scale
surveys, such as the Human Microbiome Project (10), the Earth Microbiome Project
(11), the American Gut Project (12), and Tara Oceans (13). Most DNA sequence data are
stored in either general-purpose DNA sequence repositories (e.g., NCBI SRA [14]) or
microbiome-specific databases (e.g., MG-RAST [15] and EBI Metagenomics [16]). To
support large-scale mining of the existing microbiome big data, tools have been intro-
duced to organize metagenomes with uni-
fied sequence processing standard operating
procedures (SOPs) (17), e.g., Qiita (18), gcMeta (19), and GMrepo (20). These tools typi-
cally support queries based on taxonomy terms (e.g., species name), sequence frag-
ments, or Structured Query Language (SQL)-like metadata. To support the search of
newly generated data sets against the existing microbiome big data based on taxo-
nomic or functional similarity, Microbiome Search Engine (MSE) (21) was recently
introduced, and it shows promise for search-based multiple-disease classification in a
cross-cohort, sequence-platform-insensitive, and contamination-tolerant manner
(22). However, it supports only amplicon sequencing-based data sets, which limits
the queries to those probing taxonomical similarity of microbiomes (21).

To address this limitation, here we introduce Microbiome Search Engine 2 (http://
mse.ac.cn), which enables the search of an amplicon or a shotgun WGS-based query
microbiome against a large database based on the “functional” similarity of the micro-
bioe (“taxonomical” similarity is also supported) (Fig. 1a). This platform, a significant
improvement over the previous version (21), consists of three main components
(Fig. 1b): (i) a well-maintained and regularly updated microbiome database that has
been expanding since 2016 (see Fig. S1 in the supplemental material) and currently
contains over 250,000 globally sampled (human, animal, marine, soil, etc.), curated
microbiomes (both WGS- and amplicon-based samples) that are associated with a uni-
fied scheme of metadata from 798 studies; (ii) an enhanced search engine kernel that
is compatible with both amplicon and shotgun WGS-based sequences and enables
real-time searches against the database for best matches in microbiome taxonomy or
function; and (iii) a Web-based graphical user interface that provides easy-use search-
ing, data browsing, and tutoring.

FIG 1  Design principles of MSE 2. (a) MSE 2 enables the search of a given amplicon- or shotgun-
based microbiome against a large database, based on the whole-microbiome-level taxonomical or
functional similarity between microbiomes. (b) The three key components of MSE 2 include a well-
organized database, an enhanced search engine, and a Web-based interface.
TABLE 1 Normalized metadata format of the MSE 2 database

| Metadata          | Content                                                                 |
|-------------------|-------------------------------------------------------------------------|
| Project metadata  |                                                                          |
| Project ID        | The unique identifier of this project/study in the MSE 2 database       |
| Project title     | The project name and description in the MSE 2 database                  |
| Institute         | The original producer of the study                                       |
| Principal investigator | The original principal investigator of the study                          |
| Publication title | The paper’s title of the study, if applicable                           |
| Publication journal | The journal’s name of the paper, if applicable                          |
| Source            | Link to the data source page, if applicable                             |
| Sequence type     | Sequence type, 16S rRNA gene amplicon, and/or WGS                       |
| Date              | Publishing date of the study/paper                                      |
| Sample metadata   |                                                                          |
| Sample ID         | The unique ID of this sample in the MSE 2 database, initialed by its project ID |
| Habitat domain    | Unified into 3 categories: human associated, animal associated, and environment |
| Habitat type      | Unified into 22 categories (Table S1) to further explain the habitat domain |
| Habitat details   | Detailed information of the habitat                                      |
| Sampling site     | Detailed sampling site information                                       |
| Sampling product  | The sampling material and product                                        |
| Date              | Sampling/publishing date of the sample                                  |
| Country/region    | Sampling country/region of the sample                                   |
| Gender            | Host gender of a human-associated habitat sample, if applicable         |
| Age               | Host age of a human-associated habitat sample, if applicable            |
| Description       | Additional description of sampling information, e.g., host health status, etc. |
| Amplicon sequence type | Amplicon marker and region, if applicable, e.g., 16S V4                |
| Amplicon sequencing platform | Sequencing platform for amplicon sequences, if applicable            |
| WGS               | If the sample has WGS shotgun sequences                                 |
| WGS platform      | Sequencing platform for WGS, if applicable                               |
| Function          | If the sample has functional annotation                                |
| NSTI              | The nearest sequenced taxon index to quantify the accuracy of function profiles predicted from amplicons |

RESULTS

Microbiome database. (i) Data collection and curation. Clean sequences (refer to Materials and Methods) and their metadata were collected mainly from the Qiita (18), EBI (16), SRA (14), and MG-RAST (15) repositories. The common items of metadata for each study (e.g., project name, description, publication, etc.) (Table 1, project metadata) and sample (e.g., habitat, sequencing type, sampling year, etc.) (Table 1, sample metadata) were selected and manually integrated into a specific format, while the complete original metadata were also preserved. To ensure technical comparability and searchability among microbiome samples, sequences were preprocessed and profiled by unified methods according to sequence types (i.e., amplicon based or shotgun WGS based [Table 2]; for details, see Materials and Methods).

(ii) Database statistics. After the data preprocessing and curation (details are in Materials and Methods), a total of 250,273 microbiome samples from 798 projects/studies were included in the current MSE 2 database, including 14,957 shotgun WGS-based metagenomes and 235,334 16S rRNA gene amplicons. In terms of sampling source distribution (Fig. 2), human-associated habitats are the most frequent (52.8% in total; gut, 34.2%; skin, 9.1%; oral, 6.4%, etc.), followed by animal-associated habitats.

TABLE 2 Configuration of MSE 2 for each sequence type and recommended software for preprocessing

| Search type    | Applicable sequence type                             | Recommended software(s) for sequence preprocessing (reference) | Similarity metrics |
|----------------|------------------------------------------------------|-----------------------------------------------------------------|-------------------|
| By OTU         | 16S rRNA gene amplicon                               | Parallel-META 3 (29), QIIME (35)                                | Meta-Storms (33)  |
| By species     | Shotgun WGS                                          | MetaPhAn 2 (25)                                                  | Dynamic Meta-Storms (34) |
| By function    | 16S rRNA gene amplicon and shotgun WGS             | Parallel-META 3 for a 16S rRNA gene amplicon, integrated with a C++ implantation of PICRUSt 2 (32); HUMAnN 2 for shotgun WGS (26) | Bray-Curtis        |
(23.7%), soil (6.4%), indoor environments (5.7%), and marine environments (2.7%) (for details, see Table S1 in the supplemental material).

(iii) Database organization and management. All microbiome samples are organized into two dimensions (Fig. 3). For Web-based data browsing (refer to the “Data browsing and download” section below), samples are arranged by studies and can be selected and filtered by the various metadata (e.g., habitat, sequence type, year, etc.). For searches based on taxonomical or functional similarity with microbiomes, samples were presorted by compositional features (e.g., operational taxonomic unit [OTU], species, or KEGG Orthology [KO] identifier [ID]) for indexing and searching (refer to the “Enhanced microbiome search engine” section below for details).

FIG 2 Distribution of sampling source in the microbiome database of MSE 2.

FIG 3 Database organization and structure in MSE 2. All microbiome samples were organized in two dimensions. For Web-based data browsing, samples were arranged by their studies and can be filtered via the various metadata. For whole-microbiome-level searches, samples were organized by compositional features (of either taxonomical or functional [Func.] profiles).
Enhanced microbiome search engine. (i) Whole-microbiome-level search. The search engine, as the kernel of MSE 2, was developed by C++ and optimized by OpenMP-based parallel computing. With a given query microbiome, MSE 2 searches it against the entire microbiome database for best-matched samples that have the highest taxonomical or functional similarity. The search results present the taxonomical or functional profiles of the matches, the quantitative similarity values (refer to Materials and Methods for more details) compared to the query, and their metadata information (refer to the “Microbiome search and interpretation of the search results” section below for more details). Compared to the previous version (21), which accepts only OTUs from 16S rRNA gene amplicons as the query, the search engine extended its capability by supporting OTU-based (via profiles derived from 16S rRNA gene amplicons), species-based (via profiles derived from shotgun WGS) searches, and metabolic function-based (via profiles derived from either shotgun WGS or 16S rRNA gene amplicons) searches (Fig. 4a).

(ii) Speed and scheduling. Benefited by a two-tier indexing and searching strategy (Fig. 4b; refer to Materials and Methods for details), this search engine is typically 1 to 2 orders of magnitude faster than exhaustive searches that directly compare the query to all the database samples. To test the indexing efficiency and searching speed of MSE 2, we performed OTU-based, species-based, and function-based searches against the entire database and compared the search time to that of an index-disabled exhaustive search (the exhaustive search is for in-house performance evaluation only and not provided in the public online service of MSE 2). Each process was repeated 10 times, and only the search running times (excluding the upload time, visualization time, and...
Web page loading time to avoid potential bias caused by system and network latency) were recoded and compared. The results showed that the indexing strategy accelerates the search speeds by up to 193 times, 15 times, and 605 times, respectively, for OTU-based, species-based, and KO ID-based searches (Fig. 4c and Table 3), corresponding to a real-time response of within 0.5 s for a whole-microbiome-level query against the over 250,000 samples. In addition, the online search service follows the “first come, first served” principle implemented by queue-based task scheduling, so that the computing resources are utilized efficiently.

Graphical Web-based portal. (i) Web-based user interface. MSE 2 is freely accessible via http://mse.ac.cn via Web browsers. Developed by PHP and MySQL under a Linux server, this website provides a user-friendly graphical interface (Fig. 5) for searching, data browsing, and data uploading/downloading. Tutorial materials are available for users to adjust the parameters for customized functions and result interpretation. Notifications of database updates, system maintenance, and other related information are regularly published. Users can also post any questions or bugs at the Help Desk and obtain replies via e-mail.

(ii) Microbiome search and interpretation of the search results. For microbiome searches, MSE 2 accepts the compositional features of a sample (OTUs, species, or KO IDs) as queries. Notably, query microbiomes should be preprocessed from sequences into compositional features in an way identical to that used for the database samples. Table 2 summarizes the recommended software for sequence processing for each sequence type, and the detailed analytical protocol is available via the “Search” or “Help” page. To submit a search, users first choose a search type from “Search by OTU,” “Search by species,” and “Search by function,” depending on the type of query input (Fig. 5b). Then the query can be either uploaded from a tabular plain-text file or directly pasted into the text box of the Web page. Users can also specify other parameters, such as the maximum match number (the default is 10) and the cutoff similarity (the default is 0.6).

In the result page (Fig. 5c), top-matched samples from the database are listed with sample IDs, habitats, and similarity values relative to the query (nearest sequenced taxon index [NSTI] values are also provided for 16S rRNA gene-inferred functional profiles). Each sample ID is linked to its corresponding page with detailed full metadata (e.g., source study, sampling site, sequence type, etc.). The microbial compositions of the query and matched samples are visualized via both the bar chart and the Krona-based (23) interactive animation (Fig. 5d), so as to illustrate their links and distinctions in detail. Furthermore, all the above search results are packed for download in the result page for subsequent in-depth meta-analysis and data mining by users.

(iii) Data browsing and download. The MSE 2 online service provides two ways of sample browsing.

(a) Browse by project. In the project list page, samples are organized per project, and all projects are listed and sorted by project ID. Project pages can be accessed by clicking the project ID in the list or searched by metadata key words. Each project page contains the unified metadata (e.g., study title, publication, etc.) (Table 1, project metadata), original full list of metadata, links to samples in this project, and links to its data source.

(b) Browse by sample. In the sample list page, all samples are listed and sorted by sample ID, and samples can also be selected by a metadata filter for specific habitat, sequencing type, sampling year, etc. For a given sample in the database, all the unified metadata information (Table 1, sample metadata) can be displayed and the microbial taxonomy hierarchy visualized by Krona (23) by clicking on the sample ID.

**Table 3** Performance of index-based searches and index-disabled exhaustive searches

| Search type  | Index-based search time (s) | Index-disabled search time (s) | Avg speedup |
|-------------|-----------------------------|--------------------------------|-------------|
| By OTU      | 0.241 ± 0.004               | 46.524 ± 1.029                 | 193.345     |
| By species  | 0.020 ± 0.001               | 0.300 ± 0.035                  | 14.883      |
| By function | 0.101 ± 0.002               | 60.886 ± 0.189                 | 605.076     |

*Each search procedure was repeated 10 times.*
DISCUSSION

In this work, we introduce Microbiome Search Engine 2 (MSE 2), which features (i) an expanded database of over 250,000 shotgun metagenomic and 16S rRNA gene amplicon samples associated with unified metadata collected from 798 studies and (ii) an enhanced search engine for real-time and fast (~0.5 s per query) searches for best-matched microbiomes via not just taxonomic but also functional profiles. The value of a search-based strategy has been demonstrated for defining the novelty of microbiome samples (21) and for cross-cohort disease diagnosis (22, 24). By adding a function-based dimension for these and related applications, MSE 2 should accelerate large-scale mining of the ever-expanding metagenome data space.

MATERIALS AND METHODS

Sequence preprocessing of the microbiome database. For shotgun sequences, MetaPhlAn2 (25) was used for species-level bacterial taxonomy assignment, and functional profiles were analyzed by HUMANn2 (26) using “uniref90 gene families” and annotated on the basis of the KEGG Orthology (KO).
SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.1 MB.

TABLE S1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Yinghe Qiao of QIBEBT, CAS, for server system administration and hardware maintenance and the operations research community at QIBEBT, CAS, for inspiring discussions.

X.S. acknowledges the support of grants 31771463 and 32070086 from the National Natural Science Foundation of China. This project is also supported by grants from the Shandong Natural Science Foundation, China (ZR201807060158 for G.J., ZR2019BC087 for L.L., and ZR2016QZ004 for J.X.).

We have no conflicts of interest to declare.

REFERENCES

1. Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A 103:626–631. https://doi.org/10.1073/pnas.0507535103.

2. Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow CE, Sachdeva R, Jones AC, Schwalbach MS, Rose JM, Hewson I, Patel A, Sun F, Caron DA, Fuhrman JA. 2011. Marine bacterial, archaeal and protistan
association networks reveal ecological linkages. ISME J. 5:1414–1425. https://doi.org/10.1038/ismej.2011.24.

3. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chateelier E, Yao J, Wu L, Zhou J, Ni S, Liu P, Nons B, Batto JM, Kennedy SP, Leonard P, Yuan C, Ding W, Chen Y, Hu X, Zheng B, Qian G, Xu W, Ehrlich SD, Zheng S, Li L. 2014. Alterations of the human gut microbiome in liver cirrhosis. Nature 513:53–64. https://doi.org/10.1038/nature13568.

4. Halvorson JW, Brislawn CJ, Lamendella R, Vazquez-Baeza Y, Walters WA, Branner LM, D’Amato M, Bonfiglio F, McDonald D, Gonzalez A, McClure EE, Dunklebarger MF, Knight R, Jansson JK. 2017. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Biomed 2:17004. https://doi.org/10.1038/nmib.2017.4.

5. Franzena EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser H, Reinker S, Vatanen T, Hall AB, McIlvni MJ, Sauk JS, Wilson RG, Stevens BW, Scott JM, Pierce K, Deik AA, Bullock K, Imahnn F, Porter JA, Zhernakova A, Fu J, Weensma RK, Wijmenga C, Clih CB, Vlamakis H, Hutenholzer C, Xavier R. 2019. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Biolorbil 4:293–305. https:// doi.org/10.1016/j.nmib.2019.03.006.

6. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Reinker S, Vatanen T, Hall AB, McIlvni MJ, Sauk JS, Wilson RG, Stevens BW, Scott JM, Pierce K, Deik AA, Bullock K, Imahnn F, Porter JA, Zhernakova A, Fu J, Weensma RK, Wijmenga C, Clih CB, Vlamakis H, Hutenholzer C, Xavier R. 2019. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Biolorbil 4:293–305. https://doi.org/10.1016/j.nmib.2019.03.006.

7. Feng T, Yang F, Huang S, Bo CP, Xu ZZ, Amir A, Knight R, Ling QJ, Xu J. 2015. Prediction of early childhood caries via spatial-temporal variations of oral microbiota. Cell Host Microbe 18:296–306. https://doi.org/10.1016/j.chom.2015.08.005.

8. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chateelier E, Sunagawa S, Prifti E, Vieira-Silva C, Siljander-Karlsson K, Alouisi AM, Cheung RF, Shibli-Rebecq LM, Oukhsyuen BS, Viberg M, von Knab S, Moberth S, Soobhani I, Bork P. 2014. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol 10:766. https://doi.org/10.15252/msb.20145645.

9. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpineti TV, Prieto PA, Vicente D, Hoffman K, Wei SC, Jorgensen T, Levenez F, Dore J, MetaHIT consortium, Nielsen HB, Brunak S, Alakoski S, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Prifti E,
microbial genomes and microbiomes. Nucleic Acids Res 47:D666–D677. https://doi.org/10.1093/nar/gky901.

31. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814–821. https://doi.org/10.1038/nbt.2676.

32. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI. 2020. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol 38:685–688. https://doi.org/10.1038/s41587-020-0548-6.

33. Su X, Wang X, Jing G, Ning K. 2014. GPU-Meta-Storms: computing the structure similarities among massive amount of microbial community samples using GPU. Bioinformatics 30:1031–1033. https://doi.org/10.1093/bioinformatics/btu736.

34. Jing G, Zhang Y, Yang M, Liu L, Xu J, Su X. 2020. Dynamic Meta-Storms enables comprehensive taxonomic and phylogenetic comparison of shotgun metagenomes at the species level. Bioinformatics 36:2308–2310. https://doi.org/10.1093/bioinformatics/btz910.

35. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirsch M, Reeder J, Sevinsky JR, Turnbaugh P, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303.