



OPEN NIBN Japan Microbiome Database, a database for exploring the correlations between human microbiome and health

Yi-An Chen^{1✉}, Hitoshi Kawashima², Jonguk Park¹, Attayeb Mohsen^{1,16}, Koji Hosomi^{2,3}, Takashi Nakagata⁴, Haruka Murakami^{4,5}, Hinako Nanri^{2,6}, Motohiko Miyachi^{6,7}, Jun Kunisawa^{2,8,9,10,11,12,13,14} & Kenji Mizuguchi^{1,15✉}

Recent research has highlighted the substantial impact of gut microbiome on various aspects of human health, such as obesity, inflammation, infectious diseases, and cancer. As a result, gut microbiota composition is increasingly recognized as a potential health indicator and biomarker for disease. Numerous factors, including lifestyle, diet, and physical fitness, are known to shape the composition of the human microbiome. However, a significant challenge in elucidating the relationships between these factors and the gut microbiome lies in needing a comprehensive database that integrates diverse human microbiome profiles with extensive sample metadata. To address this issue, we developed an extensive human microbiome database for healthy individuals. This initiative led to the establishment of the NIBN Japan Microbiome Database (NIBN JMD), one of the largest resources of its kind, encompassing up to 1,000 metadata points and more than 2,000 microbiome samples, including data from longitudinal studies. In this article, we describe the creation and features of NIBN JMD, detailing the data collection, processing, and database implementation. NIBN JMD is publicly accessible at <https://jmd.nibn.go.jp/>.

Keywords Human microbiome, Database, Health, Nutrition

Recent studies have revealed that gut microorganisms significantly influence our health, including obesity, inflammation, infectious diseases, and cancer. The advent of next-generation sequencing technology, coupled with decreasing costs, has led to an explosion in the amount of human microbiome data available for research. Public sequence repositories such as NCBI Sequence Read Archive (SRA)¹ (together with other International Nucleotide Sequence Database Collaboration (INSDC) members; European Bioinformatics Institute (EBI) European Nucleotide Archive (ENA)², and DDBJ Sequence Read Archive (DRA)³) now host vast collections of raw microbiome sequencing data, including 16 S rRNA amplicon and shotgun metagenomic sequences.

The human microbiota composition is influenced by various factors, including lifestyle, diet, and fitness levels, necessitating a comprehensive approach to data collection and analysis⁴. To thoroughly investigate the correlations between the human microbiome and health, it is essential to have access to large sample sizes and diverse metadata. However, the metadata available in public repositories are often restricted to specific study objectives, limiting their utility for broader research purposes⁵. Simply aggregating data from these repositories

¹Artificial Intelligence Center for Health and Biomedical Research, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan. ²Microbial Research Center for Health and Medicine, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan. ³Graduate School of Veterinary Science, Osaka Metropolitan University, Osaka, Japan. ⁴Center for Physical Activity Research, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan. ⁵Faculty of Sport and Health Science, Ritsumeikan University, Kusatsu, Japan. ⁶Center for Clinical Nutrition, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan. ⁷Faculty of Sport Sciences, Waseda University, Saitama, Japan. ⁸Graduate School of Medicine, Kobe University, Kobe, Japan. ⁹Graduate School of Medicine, The University of Osaka, Osaka, Japan. ¹⁰Graduate School of Pharmaceutical Sciences, The University of Osaka, Osaka, Japan. ¹¹Graduate School of Science, The University of Osaka, Osaka, Japan. ¹²Graduate School of Dentistry, The University of Osaka, Osaka, Japan. ¹³International Vaccine Design Center, The University of Tokyo, Tokyo, Japan. ¹⁴Research Organization for Nano and Life Innovation, Waseda University, Tokyo, Japan. ¹⁵Institute for Protein Research, The University of Osaka, Osaka, Japan. ¹⁶Libyan Centre for Dental Research, Zliten, Libya. ✉email: chenyan@nibn.go.jp; kenji@protein.osaka-u.ac.jp

Sampling place	16 S			Reference
	Samples	Subjects	Accession	
Atsugi, Kanagawa	98	50	PRJDB14457	33
Minamiuonuma, Niigata	625	625	PRJDB10528	13,18
Osaka, Osaka	277	194	PRJDB10529	18,33
Suita, Osaka	37	37	PRJDB19975	
Minato, Tokyo	13	13	PRJDB20609	
Shinjuku, Tokyo	481	481	PRJDB10526 PRJDB10527	18,19,34
Nagato, Yamaguchi	85	85	PRJDB20612	
Shunan, Yamaguchi	399	325	PRJDB10530	13,18,33
Chuo, Yamanashi	258	258	PRJDB15701	20,35–37
Total	2273	2068		

Table 1. List of studies in NIBN JMD.

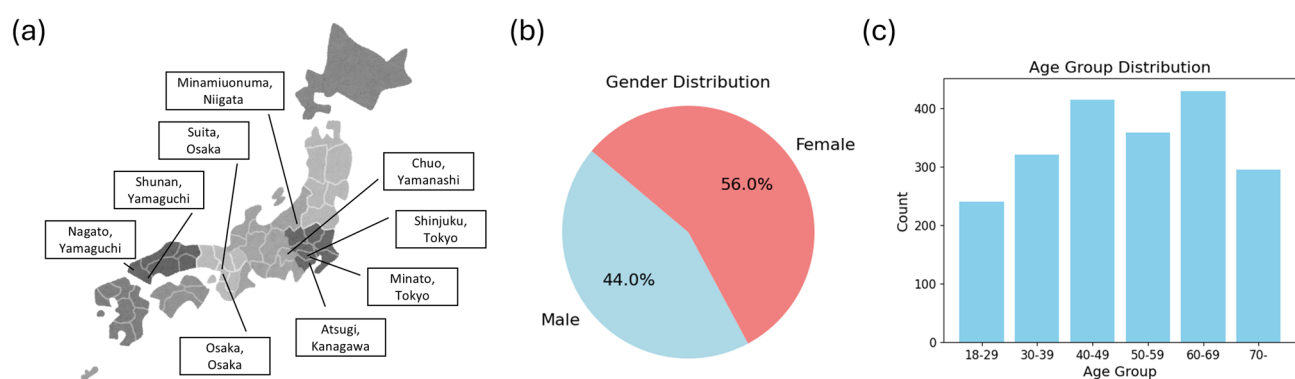


Fig. 1. The statistics of the NIBN JMD data set. (a) Shows the sampling places on the Japan map. NIBN JMD covers a wide range of areas in Japan. (b) The gender distribution in NIBN JMD. (c) The age distribution in NIBN JMD.

is insufficient due to the inconsistent and sometimes minimal metadata accompanying the microbiome samples. Additionally, variations in sample preparation and sequencing protocols across different studies can introduce significant biases and hinder comparative analyses⁶. Given those challenges, there is a critical need to develop a large-scale, standardized human microbiome database that integrates extensive sequencing data with rich metadata. Such a database would facilitate more robust and comprehensive analyses, enabling researchers to unravel the complex interactions between microbiome and human health.

Collecting human-related data presents other significant challenges. First, the costs can be substantial, particularly for advanced techniques such as metagenomic sequencing, which requires specialized equipment, reagents, and data analysis. Additionally, convincing the participants to contribute their data can take time and effort. Ethical concerns, privacy issues, and the perceived risks of data misuse often lead to hesitation or refusal, further complicating the recruitment process and limiting the scope of data collection.

In awareness of all those issues, we have been diligently working on establishing a comprehensive gut microbiome database for healthy individuals in our institute, the National Institutes of Biomedical Innovation, Health, and Nutrition (NIBN, formerly NIBIOHN), since 2015. This effort has led to the creation of the NIBN Japan Microbiome Database (NIBN JMD), which covers a vast amount of gut microbiome data derived from both 16 S rRNA gene amplicon sequencing and shotgun metagenome sequencing. Additionally, it includes extensive lifestyle-related data, such as dietary habits, nutrient intake, and physical activity levels (phenotypic metadata). Moreover, the database features the MANTA platform⁷, an open-source integrated analysis tool that facilitates comprehensive data visualization and related analyses. As we will describe in detail below, this database, by standardizing data collection, storage, and analysis protocols, would be invaluable for advancing our understanding of the microbiome's role in health and disease.

Results

The NIBN JMD database contains microbiome data from different cohort studies coordinated by NIBN. The studies were conducted in nine places across six prefectures in Japan (see Table 1; Fig. 1). Participants with no severe health issues were recruited through public websites or the local community. Their diverse geographic backgrounds enhanced the study's generalizability, and all participants provided written informed consent after thoroughly explaining the study's purposes and procedures.



Fig. 2. Screenshot of the public version of NIBN JMD. **(a)** On the entrance page. The user can select the samples of interest for analysis. There are 2,273 samples available currently. **(b)** Microbial composition in the bar chart. **(c)** Microbial composition in a heatmap with a dendrogram. **(d)** PCoA chart with envfit analysis. **(e)** Correlation analysis.

The collected fecal samples were processed using a standardized procedure, followed by data processing pipelines that adopted QIIME2⁸, Kraken 2-Bracken^{9,10}, MetaPhlAn 4¹¹, or FMAP¹² software. The microbiome profiles and cleaned-up metadata were integrated into the database using the MANTA platform⁷ and served as a web application. The Methods section describes the process in more detail.

The current NIBN JMD data set contains 2,273 samples collected from 2,068 subjects. All samples include basic metadata, such as gender and age, and depending on the cohort, some samples have detailed phenotypic metadata, including diet, nutrient intake, lifestyle, and physical activity (described in the Methods section). The subjects range in age from 18 to over 80, with 1,159 females (56.0%) and 909 males (44.0%) (Fig. 1). To ensure the participants remained anonymous, age data were processed as follows: Participants were grouped into age ranges, each containing at least five individuals. These groups were organized separately by gender and cohort, with only the age ranges disclosed. Details of the shotgun metagenome analysis results will be described elsewhere, and the full results (Kraken 2-Bracken, MetaPhlAn 4, and FMAP) will be available from a later version of the database.

NIBN JMD, powered by the MANTA platform, provides data visualization and analysis features via the web browser. Users can choose all samples or those of interest from the entrance page (Fig. 2a). The data are shown in a table and can be visualized in a bar chart (Fig. 2b) or a heatmap (Fig. 2c). Both the bar chart and the heatmap provide hierarchical clustering based on the calculated beta diversity. Some of the most widely used and essential data analysis functions are available in the system. The system precalculated and stored the distances (such as Bray-Curtis) between the sample microbiota profiles. Given those distances, the principal coordinates analysis (PCoA) projects the samples onto new axes to capture the main patterns, allowing researchers to explore sample clustering (Fig. 2d). The 'envfit' analysis function from the vegan R package¹³ was added to the PCoA plot in the latest release. The 'envfit' analysis fits the chosen variables (metadata) onto the PCoA plot to visualize how these variables influence the distribution of the samples in the plot by vectors (shown as arrows, see Fig. 2d). These vectors help researchers to interpret the strength and direction of how metadata influences the data structure.

The correlation analysis facilitates users in finding similar patterns based on the correlation coefficient (Fig. 2e). For example, the user can specify a metadata item and search for the microbiota with a correlated abundance pattern. A similar analysis can be performed for microbiota against metadata and within microbiota or metadata. The correlation analysis provides a quick view of the correlations between two items that can be a hint for the following validations. In the latest release, we added a summary table containing basic statistics for the abundance of selected microbiota or the value of the chosen metadata (Fig. 2e, above the scatter plot).

Discussion

Microbiome databases such as the Human Microbiome Project (HMP)¹⁴ and UK Biobank¹⁵ have set benchmarks for large-scale efforts to map the human microbiome across diverse populations, generating extensive genomic,

metagenomic, and phenotypic data. The HMP focuses on characterizing the human microbiome on specific body sites, while the UK Biobank integrates microbiome data with comprehensive health records from a large, predominantly European cohort. These projects are complemented by regional efforts such as Japan's ToMMO (Tohoku Medical Megabank Organization)¹⁶ and Biobank Japan¹⁷, which emphasize East Asian populations. NIBN JMD complements those resources by focusing on microbiome research, with direct access to the data and the interactive online analysis interface. In addition, with its uniform metadata and sampling and analysis protocols, NIBN JMD offers high consistency and comparability across studies, which is often a challenge in global databases where variability in data collection can hinder cross-study comparisons.

Second, by including 16 S rRNA gene amplicon and shotgun metagenomic data (currently not available yet, but more than a hundred metagenomic samples are in the process of being released), NIBN JMD provides a more comprehensive view of the microbiome, offering insights into taxonomic composition and functional potential. This dual approach is valuable for understanding how microbiomes interact with host environments. Third, including longitudinal studies in NIBN JMD, albeit limited in number currently, sets it apart by enabling researchers to track microbiome dynamics, providing critical insights into how microbiomes change in response to disease, treatment, or lifestyle changes. We are accumulating more longitudinal data with more than five repeated observations, and we plan to release some of those data after IRB approval. We hope that NIBN JMD's longitudinal component will offer a richer understanding of microbiome stability and variability than the one-time sampling from the cross-sectional study, making it a significant resource for advancing microbiome research and precision medicine.

The NIBN JMD data have been successfully applied to several studies. In addition to understanding the insightful relationship between various metadata and intestinal microorganisms, various metadata were utilized as factors in multivariate analysis to clarify the relationship in more detailed studies. For example, Park et al. identified significant associations between gut microbial variance and factors, including Bristol Stool Scale scores, gender, age, and defecation frequency¹⁸. The study showed that bowel movement frequency and specific dietary habits influence gut microbiota in Japanese populations. In another analysis, Park et al. found that vitamin B1 significantly impacts the survival and competition of gut microorganisms, particularly affecting the abundance of *Ruminococcaceae*, which require dietary vitamin B1¹⁹. The research demonstrated vitamin B1's role in butyrate production and highlighted its potential for optimizing intestinal health.

In another study, Matsuoka et al. found that barley consumption in the Japanese population is positively associated with increased abundance of *Bifidobacterium* and *Butyricoccus* in the gut²⁰. These findings suggest that barley intake influences gut microbiota composition, potentially contributing to its health benefits. A cross-sectional study of Japanese adults by Hosomi et al. identified the *Blautia* genus, particularly *B. wexlerae*, as inversely correlated with obesity and type 2 diabetes²¹. Oral administration of *B. wexlerae* in mice induced metabolic and anti-inflammatory changes, suggesting potential preventive and therapeutic strategies for metabolic disorders through unique amino acid and carbohydrate metabolism pathways.

The samples used in the studies mentioned above are mostly restricted to one or two cohorts because these studies were conducted at the initial stages of the NIBN JMD development. However, the standardized pipeline, including data collection, integration, and processing, still benefited all those studies and ensured the quality of the collected data.

NIBN JMD also provides control samples that can be used for comparison with disease patients. In the studies by Tabata et al.²² and Ikubo et al.²³, samples from NIBN JMD were used as control to compare gut microbiota between patients with atrial fibrillation (AF) or chronic thromboembolic pulmonary hypertension (CTEPH) and healthy individuals. This application highlights NIBN JMD's significance in enhancing the reliability of microbiome-based disease research by offering standardized, population-specific baseline data.

Several ongoing studies are based on the NIBN JMD data set, and we are collecting samples from other body sites, such as the oral cavity and skin. We believe that the development of NIBN JMD helps us understand the factors that influence the composition of the human microbiome and discover how this can affect our health.

The most considerable difficulty when integrating the data in NIBN JMD was the variety of metadata item names. Different systems and organizations often use varied terminologies, formats, and structures to represent the same underlying data. This disparity makes it difficult to integrate and analyze information effectively. Aware of this issue, we plan to introduce an ontology in future development. By applying an ontology, we can standardize the definitions and relationships of the data elements, promoting a common framework that all users can refer to. This helps integrate the data into our collaborations and enhances data sharing and reusing in the human microbiome research field.

The wealth of data housed in the NIBN JMD is anticipated to support a wide range of studies. Researchers can leverage this resource to gain insights into the distinctive characteristics of the Japanese population's intestinal microflora. Additionally, it will aid in exploring the intricate relationships between gut microorganisms and various health conditions or diseases, ultimately contributing to advancements in personalized medicine and public health initiatives.

Methods

Data collection

The samples are collected from cohort studies conducted by the National Institutes of Biomedical Innovation, Health, and Nutrition (NIBN) or collaborations with other research groups in Japan (Table 1). All the fecal sampling and DNA extraction follow our standardized protocols²⁴ to ensure the data quality. The extracted DNA samples were used for 16 S ribosomal DNA amplicon sequencing or shotgun metagenomic sequencing to determine the microbiome composition. All the raw sequencing data have been deposited to DDBJ Sequence Read Archive (DRA) and are available there (see Table 1 for the corresponding accession numbers).

Metadata were collected from multiple collaborators in the form of both electronic records and paper documents. The data were collected using a self-administered questionnaire, which included items on sleep patterns, bowel movement status, and medication. The paper documents were converted to electronic records with careful checks. Missing or ambiguous answers were clarified by enquiring the subjects in person. To facilitate integrating the metadata of the samples from diverse cohort studies, we predefined a set of core data items for data integration, including basic personal information, physical activities, medical history, and dietary information. The dietary information is based on the brief-type self-administered diet history questionnaire (BDHQ), a shortened version of the self-administered diet history questionnaire (DHQ)^{25,26}. A list of the covered metadata can be found in the supplementary table S1. We also maintain the list on our website [https://microbiome.nibn.go.jp/metadata_en.html] to provide up-to-date information. Given the ethics regulations, a limited number of metadata items are available on the web with no restrictions. To access other metadata items, signing a collaborative agreement with us will be required, as well as approval from the Institutional Review Board (IRB), both at the NIBN and the receiving institution.

Data processing

We used Snakemake²⁷, a Snakemake pipeline for QIIME2⁸, to process the 16 S rRNA amplicon sequencing data. The parameters used in the pipeline are as follows: bbdutk [<https://sourceforge.net/projects/bbmap/>] trimming threshold 18 for the quality trimming, forward 18 and reverse 21 for fastp²⁸ primer cropping, adopt the DADA2 algorithm²⁹, use SILVA database (version 138)³⁰ for taxonomy classification, and rarefaction to 10,000 sampling depth. The pipeline estimates the microbiota abundance and the corresponding sample diversity indices.

We also created in-house Snakemake pipelines to process the shotgun metagenomic sequencing data. These pipelines include Kraken 2⁹, followed by Bracken¹⁰ or MetaPhlAn 4¹¹ for taxonomic assignment, some additional calculations for the diversity indices, and data formatting. One more pipeline runs FMAP¹² for functional profiling. To ensure the consistency of the taxonomy across different reference databases, we annotate all the taxa using the NCBI Taxonomy identifiers (taxon IDs)³¹.

Despite our attempt to unify the metadata items, some variations remain due to the different study topics. To solve this problem, we defined a massive collection of metadata items with internal identifiers to cover the metadata from diverse cohorts. The metadata item names from each cohort were integrated by text matching and assigned internal identifiers, such as 10,010 for age and 10,012 for gender, to maintain consistency across the dataset. Data validation checks were conducted to identify and correct errors, such as text entries in numeric fields or invalid category options. After validation, the metadata were stored in NIBN JMD for further analysis.

Database development

We stored and shared the data on the MANTA platform⁷ for NIBN JMD. MANTA is an integrated analysis platform that hosts gut microbiome composition and phenotypic metadata, such as diet and exercise. It stores the data and provides data analysis functions, including data visualizations. The platform comprises a relational database (PostgreSQL) and a web application as the user interface. We reimplemented the whole system to improve the MANTA platform and added some additional features described in the Results section.

To facilitate continuous updating of NIBN JMD, we used an in-house script to import the data from the processed data described previously. The data importing script is followed by another script to calculate the all-against-all beta diversity³², including Bray-Curtis distance and Jaccard distance, for each rank when new samples are added. These scripts will be included in the newly implemented MANTA platform.

Data availability

Raw sequence read data are available from DRA of DDBJ under the accession numbers in Table 1. Approximately 2,000 samples, along with basic metadata and a subset of dietary and nutritional data from BDHQ, are available at <https://jmd.nibn.go.jp/>.

Received: 2 April 2025; Accepted: 26 May 2025

Published online: 04 June 2025

References

1. Leinonen, R., Sugawara, H. & Shumway, M. International nucleotide sequence database collaboration. The sequence read archive. *Nucleic Acids Res.* **39**, D19–D21. <https://doi.org/10.1093/nar/gkq1019> (2011).
2. Leinonen, R. et al. The European nucleotide archive. *Nucleic Acids Res.* **39**, D28–D31. <https://doi.org/10.1093/nar/gkq967> (2011).
3. Kaminuma, E. et al. DDBJ launches a new archive database with analytical tools for next-generation sequence data. *Nucleic Acids Res.* **38**, D33–D38. <https://doi.org/10.1093/nar/gkp847> (2010).
4. Gilbert, J. A. et al. Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400. <https://doi.org/10.1038/nm.4517> (2018).
5. Kasmanas, J. C. et al. HumanMetagenomeDB: a public repository of curated and standardized metadata for human metagenomes. *Nucleic Acids Res.* **49**, D743–D750. <https://doi.org/10.1093/nar/gkaa1031> (2021).
6. Sinha, R. et al. Assessment of variation in microbial community amplicon sequencing by the microbiome quality control (MBQC) project consortium. *Nat. Biotechnol.* **35**, 1077–1086. <https://doi.org/10.1038/nbt.3981> (2017).
7. Chen, Y. A. et al. MANTA, an integrative database and analysis platform that relates microbiome and phenotypic data. *PLoS One*. **15**, e0243609. <https://doi.org/10.1371/journal.pone.0243609> (2020).
8. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857. <https://doi.org/10.1038/s41587-019-0209-9> (2019).
9. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with kraken 2. *Genome Biol.* **20** <https://doi.org/10.1186/s13059-019-1891-0> (2019).
10. Lu, J., Breitwieser, F. P., Thielen, P., Salzberg, & Bracken S. L. Estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104. <https://doi.org/10.7717/peerj-cs.104> (2017).

11. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using metaphlan 4. *Nat. Biotechnol.* **41**, 1633–1644. <https://doi.org/10.1038/s41587-023-01688-w> (2023).
12. Kim, J. et al. Functional mapping and analysis pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinform.* **17**, <https://doi.org/10.1186/s12859-016-1278-0> (2016).
13. Oksanen, J. et al. vegan: Community Ecology Package. R package version 2.6-4. <https://vegandevs.github.io/vegan/>.
14. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221. <https://doi.org/10.1038/nature11209> (2012).
15. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779. <https://doi.org/10.1371/journal.pmed.1001779> (2015).
16. Fuse, N. et al. Establishment of integrated biobank for precision medicine and personalized healthcare: the Tohoku medical megabank project. *JMA J.* **2**, 113–122. <https://doi.org/10.31662/jmaj.2019-0014> (2019).
17. Nagai, A. et al. Overview of the biobank Japan project: study design and profile. *J. Epidemiol.* **27**, S2–S8. <https://doi.org/10.1016/j.je.2016.12.005> (2017).
18. Park, J. et al. Comprehensive analysis of gut microbiota of a healthy population and covariates affecting microbial variation in two large Japanese cohorts. *BMC Microbiol.* **21**(151). <https://doi.org/10.1186/s12866-021-02215-0> (2021).
19. Park, J. et al. Dietary vitamin B1 intake influences gut microbial community and the consequent production of short-chain fatty acids. *Nutrients* **14**, 2078. <https://doi.org/10.3390/nu14102078> (2022).
20. Matsuoka, T. et al. Relationships between barley consumption and gut microbiome characteristics in a healthy Japanese population: a cross-sectional study. *BMC Nutr.* **8**(23). <https://doi.org/10.1186/s40795-022-00500-3> (2022).
21. Hosomi, K. et al. Oral administration of *Blautia wexlerae* ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. *Nat. Commun.* **13**, 4477. <https://doi.org/10.1038/s41467-022-32015-7> (2022).
22. Tabata, T. et al. Gut microbial composition in patients with atrial fibrillation: effects of diet and drugs. *Heart Vessels.* **36**, 105–114 (2021).
23. Ikubo, Y. et al. Altered gut microbiota and its association with inflammation in patients with chronic thromboembolic pulmonary hypertension: a single-center observational study in Japan. *BMC Pulm Med.* **22**, 138. <https://doi.org/10.1186/s12890-022-01932-0> (2022).
24. Hosomi, K. et al. Method for preparing DNA from feces in guanidine thiocyanate solution affects 16S rRNA-based profiling of human microbiota diversity. *Sci. Rep.* **7**, 4339. <https://doi.org/10.1038/s41598-017-04511-0> (2017).
25. Sasaki, S., Yanagibori, R. & Amano, K. Self-administered diet history questionnaire developed for health education: a relative validation of the test-version by comparison with 3-day diet record in women. *J. Epidemiol.* **8**, 203–215. <https://doi.org/10.2188/jea.8.203> (1998).
26. Kobayashi, S. et al. Comparison of relative validity of food group intakes estimated by comprehensive and brief-type self-administered diet history questionnaires against 16 d dietary records in Japanese adults. *Public. Health Nutr.* **14**, 1200–1211. <https://doi.org/10.1017/S1368980011000504> (2011).
27. Mohsen, A. et al. A dynamic snakemake pipeline for microbiome data analysis with QIIME2. *Front. Bioinform.* **2**, 893933. <https://doi.org/10.3389/fbinf.2022.893933> (2022).
28. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560> (2018).
29. Callahan, B. J. et al. DADA2: high-resolution sample inference from illumina amplicon data. *Nat. Methods.* **13**, 581–583. <https://doi.org/10.1038/nmeth.3869> (2016).
30. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596. <https://doi.org/10.1093/nar/gks1219> (2013).
31. Schoch, C. L. et al. NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database.* **2020**, 1–21. <https://doi.org/10.1093/database/baaa062> (2020).
32. By IMPACTT investigators. Beta-diversity distance matrices for microbiome sample size and power calculations – How to obtain good estimates. *Comput. Struct. Biotechnol. J.* **20**, 2259–2267. <https://doi.org/10.1016/j.csbj.2022.04.032> (2022).
33. Sawane, K. et al. Identification of human gut microbiome associated with enterolignan production. *Microorganisms* **10**, 2169. <https://doi.org/10.3390/microorganisms10112169> (2022).
34. Park, J. et al. Ramen consumption and gut microbiota diversity in Japanese women: cross-sectional data from the NEXIS cohort study. *Microorganisms* **11**, 1892. <https://doi.org/10.3390/microorganisms11081892> (2023).
35. Maruyama, S. et al. Classification of the occurrence of dyslipidemia based on gut bacteria related to barley intake. *Front. Nutr.* **9**, 812469. <https://doi.org/10.3389/fnut.2022.812469> (2022).
36. Maruyama, S. et al. Characteristic gut bacteria in high barley consuming Japanese individuals without hypertension. *Microorganisms* **11**, 1246. <https://doi.org/10.3390/microorganisms11051246> (2023).
37. Maruyama, S. et al. High barley intake in non-obese individuals is associated with high natto consumption and abundance of butyrate-producing bacteria in the gut: a cross-sectional study. *Front. Nutr.* **11**, 1434150. <https://doi.org/10.3389/fnut.2024.1434150> (2024).

Author contributions

YAC wrote the main manuscript. KH, TN, HM, and HN acquired the data. YAC, HK, JP, and AM processed and analyzed the data. YAC developed the database. MM, JK, and KM acquired the funding and conducted the project. All authors reviewed the manuscript.

Declarations

Ethics declarations

This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the National Institutes of Biomedical Innovation, Health and Nutrition (Approval number 260-05). All participants in this study provided written informed consent. The publication and the database do not contain personally identifiable information.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-04339-z>.

Correspondence and requests for materials should be addressed to Y.-A.C. or K.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025