Nucleic Acids Research, 2025, **53**, D783–D788
https://doi.org/10.1093/nar/gkae1002
Advance access publication date: 30 October 2024
Database issue

gutMGene v2.0: an updated comprehensive database for target genes of gut microbes and microbial metabolites

Changlu Qi[1], Guoyou He[1], Kai Qian[1], Siyuan Guan[1], Zhaohai Li[1], Shuang Liang[1], Juntao Liu[2], Xianzhe Ke[1], Sainan Zhang[1], Minke Lu[1], Liang Cheng[1,3,*] and Xue Zhang[3,4,5,*]

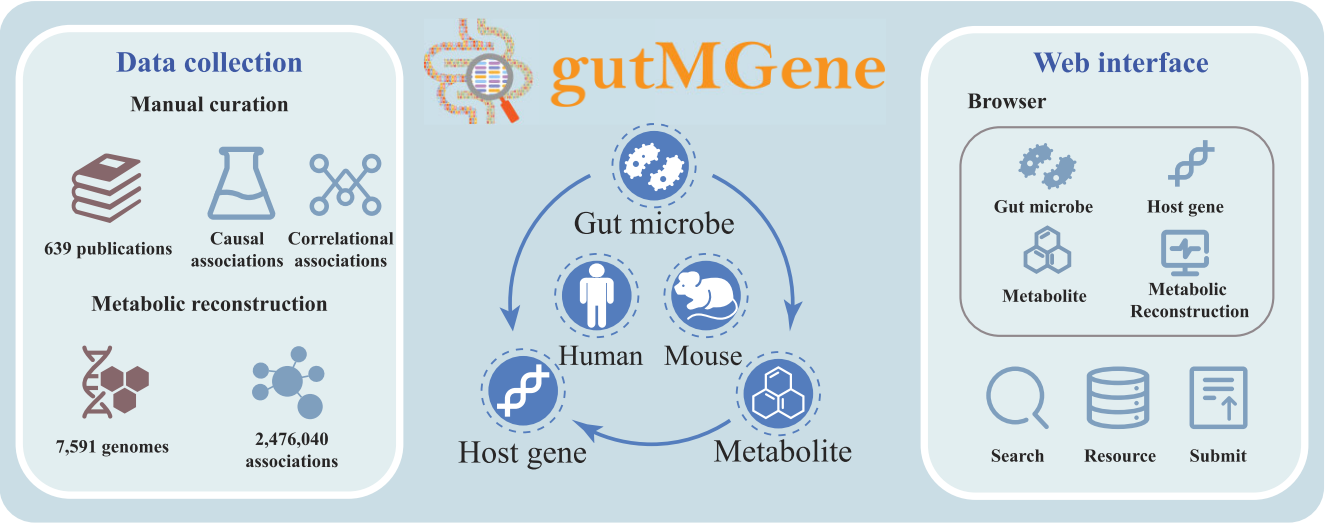
[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150081, China
[2]School of Basic Medical Sciences, Harbin Medical University, Harbin, Heilongjiang 150081, China
[3]National Health Commission (NHC) Key Laboratory of Molecular Probes and Targeted Diagnosis and Therapy, Harbin Medical University, Harbin, Heilongjiang 150081, China
[4]McKusick-Zhang Center for Genetic Medicine, State Key Laboratory of Complex Severe and Rare Diseases, Department of Medical Genetics, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing 100005, China
[5]Department of Child and Adolescent Health, School of Public Health, Harbin Medical University, Harbin, Heilongjiang 150081, China

[*]To whom correspondence should be addressed. Tel: +86 153 0361 4540; Email: liangcheng@hrbmu.edu.cn
Correspondence may also be addressed to Xue Zhang. Email: xuezhang@hrbmu.edu.cn

Abstract

The gut microbiota is essential for various physiological functions in the host, primarily through the metabolites it produces. To support researchers in uncovering how gut microbiota contributes to host homeostasis, we launched the gutMGene database in 2022. In this updated version, we conducted an extensive review of previous papers and incorporated new papers to extract associations among gut microbes, their metabolites, and host genes, carefully classifying these as causal or correlational. Additionally, we performed metabolic reconstructions for representative gut microbial genomes from both human and mouse. gutMGene *v2.0* features an upgraded web interface, providing users with improved accessibility and functionality. This upgraded version is freely available at http://bio-computing.hrbmu.edu.cn/gutmgene. We believe that this new version will greatly advance research in the gut microbiota field by offering a comprehensive resource.

Graphical abstract



Introduction

Gut microbiota colonizes the intestine and is essential for various host physiological functions, primarily through the metabolites it produces (1–3). These metabolites can directly affect the gut and also influence other organs via the bloodstream, as demonstrated in the established gut-brain axis and

Received: July 12, 2024. **Revised:** October 1, 2024. **Editorial Decision:** October 12, 2024. **Accepted:** October 17, 2024

gut-liver axis (4–8). Therefore, pinpointing specific microbes that generate particular metabolites and understanding the impact of these compounds on host physiology is crucial for utilizing gut microbes in therapeutic strategies (9–11).

The first version of gutMGene, released in 2022, involved gathering experimentally validated associations among gut microbes, their metabolites, and host genes from the scientific literature (12). The interactions among these components include both correlational and causal relationships. Distinguishing these types is vital for accurate interpretation of results and for avoiding misleading conclusions that may arise from conflating correlation with causation. Furthermore, to accurately characterize the taxonomy and function of microbial ecosystems, microbial reference genomes are continually refined and expanded. Almeida *et al*. compiled a substantial collection of prokaryotic genomes, leading to the development of the Unified Human Gastrointestinal Genome collection, which encompasses 4 644 representative gut prokaryotic species (13). Likewise, Huang *et al*. created the Vaginal Microbial Genome Collection, which includes 786 prokaryotic species, 11 fungal species and 4263 viral operational taxonomic units (14). Simultaneously, this ongoing enhancement has established genome-based metabolic reconstruction as a powerful approach for studying gut microbial metabolism. For instance, DEMETER (15) employs comparative genomics with the PubSEED platform (16) to annotate metabolic functions in bacterial and archaeal strains. It uses gap-filling techniques based on the bidirectional best-hit method, assesses pathway completeness, and incorporates phylogenetic analysis to annotate drug metabolism genes. CarveMe (17) utilizes a manually curated universal bacterial model from the BiGG database (18), integrating specialized biomass templates and gene annotations via BLAST searches (19). It employs mixed integer linear programming for model carving, enabling the creation of organism-specific and microbial community models, as well as gap-filling and experimental constraint incorporation. This approach enhances the prediction of metabolic capabilities and gene essentiality under various conditions. metaGEM (20) is an end-to-end pipeline that begins with metagenomic data assembly and employs CarveMe for metabolic reconstruction. The MIGRENE toolbox (21) generates species-specific genome-scale metabolic models by integrating a non-redundant microbial gene catalog with a metabolic model, establishing reaction profiles and scores for each metagenome species using a reference model and taxonomic data. This toolbox facilitates individualized metabolic microbiome analysis, producing outputs such as reactobiomes, reaction abundance, and community modeling. gapseq (22) is a novel software designed for pathway analysis and metabolic network reconstruction, leveraging multiple biochemistry databases to predict pathways and key enzymes. It constructs genome-scale metabolic models using a curated reaction database and employs a linear programming-based gap-filling algorithm to identify and resolve metabolic gaps, enhancing biomass formation and metabolic function. KBase (23) utilizes the ModelSEED pipeline (24), integrating genomic annotations with a comprehensive biochemical database to generate genome-scale metabolic models. These models are constructed through gap-filling algorithms that resolve network inconsistencies and enable functional flux-balance analysis for predicting microbial growth and metabolism. However, utilizing these analytical tools often requires substantial time investment and a certain level of programming proficiency. Intuitively presenting microbial metabolites derived from genome-based metabolic reconstructions significantly enhances research convenience. Existing databases, such as VMH (25), currently contain only a limited number of genomes, which is insufficient to meet the growing research demands.

Therefore, to update gutMGene, a thorough review of all literature referenced in v1.0 was conducted, alongside the integration of new studies to classify relationships among gut microbes, metabolites and host genes, distinctly separating causal from correlational links. These associations were further divided into three categories: gut microbe–metabolite, microbial metabolite–host gene and gut microbe–host gene relationships, while also incorporating different strains of the same species. Metabolic reconstructions were carried out for 4744 human and 2847 mouse gut microbial reference genomes, respectively. Lastly, the system interface was enhanced with a minimalist and aesthetically pleasing design to improve user experience.

Data collection and database content

The process of updating gutMGene involved a comprehensive review of all literature cited in version 1.0, supplemented by new studies published from 31 October 2021 to 31 October 2023. A search of the PubMed database was conducted using key terms such as ‘gut’, ‘intestinal’, ‘microbiota’, ‘microbiome’, ‘metabolite’ and ‘gene’. Each downloaded paper was carefully examined to identify meaningful associations among gut microbe–metabolite, metabolite–host gene and gut microbe–host gene pairs, with irrelevant studies being excluded. These identified associations were categorized as either causal or correlational. Causal associations stem from controlled experiments that manipulate a specific variable—either a gut microbe or a microbial metabolite—to observe resulting changes in metabolites or host genes. Conversely, correlational associations are derived from statistical correlation analyses, including Pearson correlation methods. To maintain consistency across the database, the nomenclature for gut microbes was standardized using the NCBI taxonomy database (26), with their corresponding IDs documented. Metabolite names were aligned with PubChem (27), and IDs were sourced from HMDB (28), ChEBI (29), KEGG (30), fooDB (https://foodb.ca/) and Metabolights (31). Gene names followed the standards set by the NCBI Gene database (32), ensuring accurate symbol usage and ID recording. Additional information regarding metabolite substrates, sample types, experimental methodologies, measurement techniques, and descriptive contexts was extracted from the articles. The distribution of causal and correlational associations for both human and mouse models is illustrated in Figure 1A and B. The current version of gutMGene now includes 1338 curated associations among 282 gut microbes, 278 microbial metabolites, and 238 host genes in Human, and 3522 curated associations among 341 gut microbes, 501 microbial metabolites and 609 host genes in mouse.

For the metabolic reconstruction effort, 4744 representative human gut microbial genomes (v2.0.2) and 2847 representative mouse gut microbial genomes (v1.0), along with their quality metrics, were obtained from MGnify (33). These genomes were meticulously selected based on rigorous criteria, ensuring they met standards for medium to high quality, defined as over 50% completeness and <5% contamination. To facilitate further analysis, genome annotations were per-

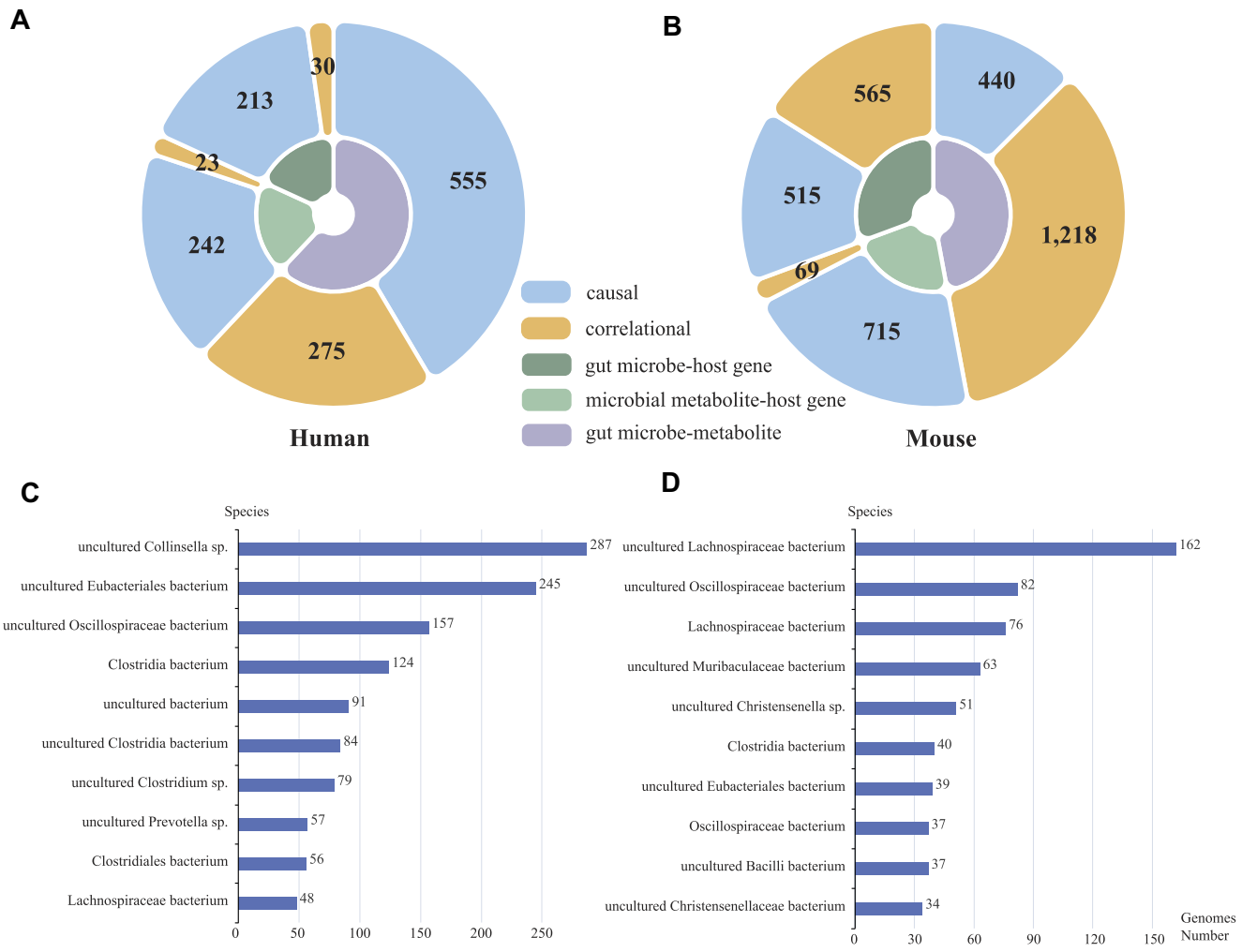


Figure 1. Statistics describing the data in gutMGene *v2.0*. (**A**) The distribution of causal and correlational associations in Human. (**B**) The distribution of causal and correlational associations in Mouse. (**C**) The top 10 species by genome count in metabolic reconstruction in human. (**D**) The top 10 species by genome count in metabolic reconstruction in Mouse.

formed Prokka (34), resulting in GBK files that were subsequently uploaded to KBase. The 'Build Multiple Metabolic Models' application was employed to generate draft metabolic reconstructions, which were further refined using the DEMETER pipeline. Information pertaining to metabolites was extracted from the generated .mat files, while taxonomic annotations for all genomes were provided through the GTDB (release 214) (35). This comprehensive compilation ultimately resulted in the identification of 1 554 878 associations in Human and 92,1162 associations in Mouse. A comparative overview of the two data versions is presented in Table 1. Figure 1C and D illustrate the top 10 species by genome count within the metabolic reconstructions in Human and Mouse, highlighting substantial differences. Furthermore, the annotation of multiple genomes to the same species underscores the necessity for researchers to delve deeper into these genomes, facilitating the exploration of nuanced difference.

Database access

gutMGene v2.0 is publicly available at http://bio-computing.hrbmu.edu.cn/gutmgene. This resource allows users to navigate, search, and retrieve associations involving gut microbes, metabolites, and host genes through a user-friendly interface. Navigation is streamlined via hyperlinks positioned on the right side of the **Home** page or within the **Browse** dropdown menu. For example, selecting 'Gut Microbe' enables exploration of metabolites synthesized by specific microbes or the host genes they modulate. Users may also switch between various host species or association categories by utilizing the designated tabs. The 'Evidence' column denotes whether an association is characterized as correlational or causal, while the 'Evidence Number' column indicates the count of publications documenting each association. Detailed evidence can be accessed by clicking the arrow in the 'Details' column (Figure 2). Additionally, the 'Metabolic Reconstruction' section provides insights into microbial genomic information, with metabolite specifics accessible through the same 'Details' feature. On the **Search** page, users can input relevant gut microbes, metabolites, or host genes, leveraging autocomplete suggestions to refine their queries before submission. The ChEBI ontology (36) further enhances search capabilities by allowing exploration of broader chemical categories. The **Resource** page grants comprehensive access to all genomes and metabolites linked to metabolic reconstructions. Furthermore, the gutM-

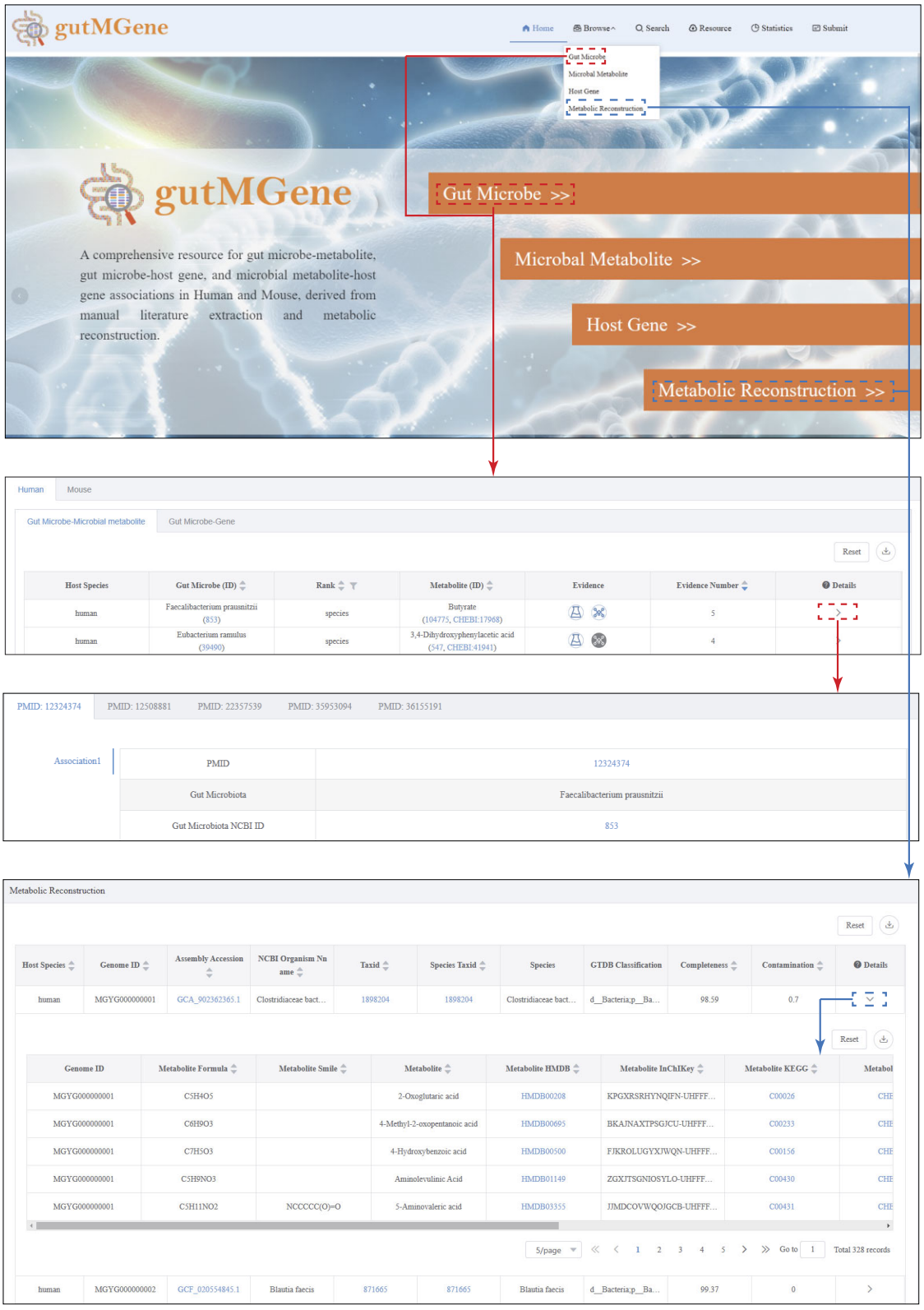


Figure 2. Schematic workflow of browsing gutMGene *v2.0*.

Table 1. The number of gut microbes, microbial metabolites, host genes and their associations in human and mouse

Version	Host species	Associations source	No. of gut microbes	No. of microbial metabolites	No. of host genes	No. of associations (gut microbe-metabolite; microbial metabolite-gene; gut microbe–host gene)
1.0	Human	Literature-based associations	193	203	207	532; 233; 182
1.0	Mouse	Literature-based associations	120	144	446	359; 512; 317
2.0	Human	Literature-based associations	282	277	238	830; 265; 243
2.0	Mouse	Literature-based associations	341	501	609	1658; 784; 1 080
2.0	Human	Metabolic reconstitution-based associations	4744	611	-	1554 878; -; -
2.0	Mouse	Metabolic reconstitution-based associations	2847	583	-	921 162; -; -

Gene v2.0 web server features a **Submit** page, enabling researchers to input newly validated experimental associations into the database.

Conclusion

The gut microbiota is vital for regulating host physiological functions through its metabolites, highlighting the need for a comprehensive exploration of microbial mechanisms. In this updated version of gutMGene, all associations are classified as causal or correlational, providing clarity for researchers. The dataset has been enriched not only by extracting additional gut microbe–metabolite, microbial metabolite–host gene and gut microbe–host gene associations from a wide range of literature, but also by conducting metabolic reconstructions of representative gut microbial genomes. With an upgraded system interface, gutMGene v2.0 now includes 4860 literature-based associations and 2 476 040 associations derived from metabolic reconstructions. This resource will continue to be updated, enhancing our understanding of the intricate interactions between gut microbes and host physiology.

Funding

National Natural Science Foundation of China [62222104, 62172130]; Heilongjiang Postdoctoral Fund [LBH-Q20030]. Funding for open access charge: National Natural Science Foundation of China [62222104, 62172130]; Heilongjiang Postdoctoral Fund [LBH-Q20030].

Conflict of interest statement

None declared.

References

1. Lee,J.Y., Tsolis,R.M. and Baumler,A.J. (2022) The microbiome and gut homeostasis. *Science*, **377**, eabp9960.
2. Lavelle,A. and Sokol,H. (2020) Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.*, **17**, 223–237.
3. Fan,Y. and Pedersen,O. (2021) Gut microbiota in human metabolic health and disease. *Nat. Rev. Micro.*, **19**, 55–71.
4. Wikoff,W.R., Anfora,A.T., Liu,J., Schultz,P.G., Lesley,S.A., Peters,E.C. and Siuzdak,G. (2009) Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 3698–3703.
5. Wang,Z., Wang,Z., Lu,T., Chen,W., Yan,W., Yuan,K., Shi,L., Liu,X., Zhou,X., Shi,J., *et al.* (2022) The microbiota-gut-brain axis in sleep disorders. *Sleep Med. Rev.*, **65**, 101691.
6. Qi,C., Wang,P., Fu,T., Lu,M., Cai,Y., Chen,X. and Cheng,L. (2021) A comprehensive review for gut microbes: technologies, interventions, metabolites and diseases. *Brief. Funct. Genomics*, **20**, 42–60.
7. Hsu,C.L. and Schnabl,B. (2023) The gut-liver axis and gut microbiota in health and liver disease. *Nat. Rev. Micro.*, **21**, 719–733.
8. Dinan,T.G. and Cryan,J.F. (2017) Gut-brain axis in 2016: brain-gut-microbiota axis - mood, metabolism and behaviour. *Nat. Rev. Gastroenterol. Hepatol.*, **14**, 69–70.
9. Lee,M. and Chang,E.B. (2021) Inflammatory bowel diseases (IBD) and the microbiome-searching the crime scene for clues. *Gastroenterology*, **160**, 524–537.
10. Kujawa,D., Laczmanski,L., Budrewicz,S., Pokryszko-Dragan,A. and Podbielska,M. (2023) Targeting gut microbiota: new therapeutic opportunities in multiple sclerosis. *Gut Microbes*, **15**, 2274126.
11. Fernandes,M.R., Aggarwal,P., Costa,R.G.F., Cole,A.M. and Trinchieri,G. (2022) Targeting the gut microbiota for cancer therapy. *Nat. Rev. Cancer*, **22**, 703–722.
12. Cheng,L., Qi,C., Yang,H., Lu,M., Cai,Y., Fu,T., Ren,J., Jin,Q. and Zhang,X. (2022) gutMGene: a comprehensive database for target genes of gut microbes and microbial metabolites. *Nucleic Acids Res.*, **50**, D795–D800.
13. Almeida,A., Nayfach,S., Boland,M., Strozzi,F., Beracochea,M., Shi,Z.J., Pollard,K.S., Sakharova,E., Parks,D.H., Hugenholtz,P., *et al.* (2021) A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.*, **39**, 105–114.
14. Huang,L., Guo,R., Li,S., Wu,X., Zhang,Y., Guo,S., Lv,Y., Xiao,Z., Kang,J., Meng,J., *et al.* (2024) A multi-kingdom collection of 33,804 reference genomes for the human vaginal microbiome. *Nat. Biotechnol.*, **9**, 2185–2200.
15. Heinken,A., Hertel,J., Acharya,G., Ravcheev,D.A., Nyga,M., Okpala,O.E., Hogan,M., Magnusdottir,S., Martinelli,F., Nap,B., *et al.* (2023) Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. *Nat. Biotechnol.*, **41**, 1320–1331.
16. Overbeek,R., Olson,R., Pusch,G.D., Olsen,G.J., Davis,J.J., Disz,T., Edwards,R.A., Gerdes,S., Parrello,B., Shukla,M., *et al.* (2014) The

SEED and the rapid annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.*, **42**, D206–D214.
17. Machado,D., Andrejev,S., Tramontano,M. and Patil,K.R. (2018) Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.*, **46**, 7542–7553.
18. Norsigian,C.J., Pusarla,N., McConn,J.L., Yurkovich,J.T., Drager,A., Palsson,B.O. and King,Z. (2020) BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.*, **48**, D402–D406.
19. Camacho,C., Boratyn,G.M., Joukov,V., Vera Alvarez,R. and Madden,T.L. (2023) ElasticBLAST: accelerating sequence search via cloud computing. *BMC Bioinformatics*, **24**, 117.
20. Zorrilla,F., Buric,F., Patil,K.R. and Zelezniak,A. (2021) metaGEM: reconstruction of genome scale metabolic models directly from metagenomes. *Nucleic Acids Res.*, **49**, e126.
21. Bidkhori,G. and Shoaie,S. (2024) MIGRENE: the toolbox for microbial and individualized GEMs, reactobiome and community network modelling. *Metabolites*, **14**, 132.
22. Zimmermann,J., Kaleta,C. and Waschina,S. (2021) gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol.*, **22**, 81.
23. Arkin,A.P., Cottingham,R.W., Henry,C.S., Harris,N.L., Stevens,R.L., Maslov,S., Dehal,P., Ware,D., Perez,F., Canon,S., *et al.* (2018) KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.*, **36**, 566–569.
24. Henry,C.S., DeJongh,M., Best,A.A., Frybarger,P.M., Linsay,B. and Stevens,R.L. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.*, **28**, 977–982.
25. Noronha,A., Modamio,J., Jarosz,Y., Guerard,E., Sompairac,N., Preciat,G., Danielsdottir,A.D., Krecke,M., Merten,D., Haraldsdottir,H.S., *et al.* (2019) The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.*, **47**, D614–D624.
26. Schoch,C.L., Ciufo,S., Domrachev,M., Hotton,C.L., Kannan,S., Khovanskaya,R., Leipe,D., Mcveigh,R., O'Neill,K., Robbertse,B., *et al.* (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)*, **2020**, baaa062.
27. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B., *et al.* (2021) PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.*, **49**, D1388–D1395.
28. Wishart,D.S., Guo,A., Oler,E., Wang,F., Anjum,A., Peters,H., Dizon,R., Sayeeda,Z., Tian,S., Lee,B.L., *et al.* (2022) HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res.*, **50**, D622–D631.
29. Hastings,J., Owen,G., Dekker,A., Ennis,M., Kale,N., Muthukrishnan,V., Turner,S., Swainston,N., Mendes,P. and Steinbeck,C. (2016) ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res.*, **44**, D1214–D1219.
30. Jin,Z., Sato,Y., Kawashima,M. and Kanehisa,M. (2023) KEGG tools for classification and analysis of viral proteins. *Protein Sci.*, **32**, e4820.
31. Yurekten,O., Payne,T., Tejera,N., Amaladoss,F.X., Martin,C., Williams,M. and O'Donovan,C. (2024) MetaboLights: open data repository for metabolomics. *Nucleic Acids Res.*, **52**, D640–D646.
32. Brown,G.R., Hem,V., Katz,K.S., Ovetsky,M., Wallin,C., Ermolaeva,O., Tolstoy,I., Tatusova,T., Pruitt,K.D., Maglott,D.R., *et al.* (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–D42.
33. Richardson,L., Allen,B., Baldi,G., Beracochea,M., Bileschi,M.L., Burdett,T., Burgin,J., Caballero-Perez,J., Cochrane,G., Colwell,L.J., *et al.* (2023) MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res.*, **51**, D753–D759.
34. Seemann,T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**, 2068–2069.
35. Parks,D.H., Chuvochina,M., Rinke,C., Mussig,A.J., Chaumeil,P.A. and Hugenholtz,P. (2022) GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.*, **50**, D785–D794.
36. Hastings,J., Owen,G., Dekker,A., Ennis,M., Kale,N., Muthukrishnan,V., Turner,S., Swainston,N., Mendes,P. and Steinbeck,C. (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–D1219.

Received: July 12, 2024. **Revised:** October 1, 2024. **Editorial Decision:** October 12, 2024. **Accepted:** October 17, 2024