Nucleic Acids Research, 2024, **52**, D579–D585
https://doi.org/10.1093/nar/gkad1086
Advance access publication date: 22 November 2023
Database issue

ABC-HuMi: the Atlas of Biosynthetic Gene Clusters in the Human Microbiome

Pascal Hirsch [1,†], Azat Tagirdzhanov [1,2,†], Aleksandra Kushnareva [1,2], Ilia Olkhovskii [1,2,3], Simon Graf [4], Georges P. Schmartz [1], Julian D. Hegemann [2,5], Kenan A.J. Bozhüyük [2], Rolf Müller [2,5,*], Andreas Keller [1,2,*] and Alexey Gurevich [1,2,4,*]

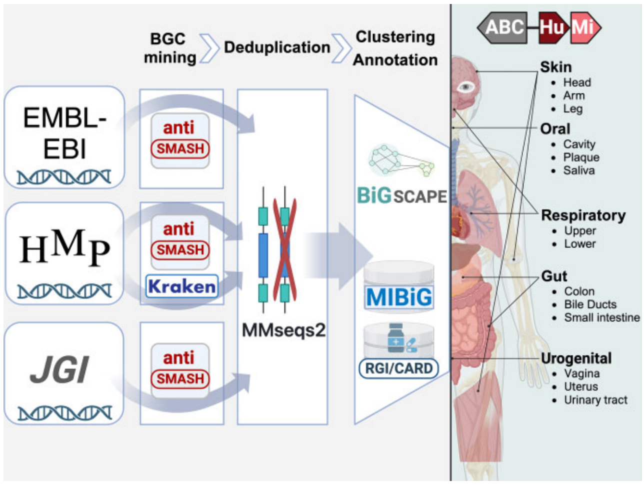
[1] Center for Bioinformatics, Saarland University, Saarbrücken 66123, Germany
[2] Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarbrücken 66123, Germany
[3] Saarbrücken Graduate School of Computer Science, Saarland University, Saarbrücken 66123, Germany
[4] Department of Computer Science, Saarland University, Saarbrücken 66123, Germany
[5] Department of Pharmacy, Saarland University, Saarbrücken 66123, Germany

[*] To whom correspondence should be addressed. Tel: +49 681 988063002; Fax: +49 681 988066001; Email: rolf.mueller@helmholtz-hips.de
Correspondence may also be addressed to Andreas Keller. Email: andreas.keller@ccb.uni-saarland.de
Correspondence may also be addressed to Alexey Gurevich. Email: alexey.gurevich@helmholtz-hips.de
[†] The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Abstract

The human microbiome has emerged as a rich source of diverse and bioactive natural products, harboring immense potential for therapeutic applications. To facilitate systematic exploration and analysis of its biosynthetic landscape, we present ABC-HuMi: the Atlas of Biosynthetic Gene Clusters (BGCs) in the Human Microbiome. ABC-HuMi integrates data from major human microbiome sequence databases and provides an expansive repository of BGCs compared to the limited coverage offered by existing resources. Employing state-of-the-art BGC prediction and analysis tools, our database ensures accurate annotation and enhanced prediction capabilities. ABC-HuMi empowers researchers with advanced browsing, filtering, and search functionality, enabling efficient exploration of the resource. At present, ABC-HuMi boasts a catalog of 19 218 representative BGCs derived from the human gut, oral, skin, respiratory and urogenital systems. By capturing the intricate biosynthetic potential across diverse human body sites, our database fosters profound insights into the molecular repertoire encoded within the human microbiome and offers a comprehensive resource for the discovery and characterization of novel bioactive compounds. The database is freely accessible at https://www.ccb.uni-saarland.de/abc_humi/.

Graphical abstract



Introduction

Bioactive compounds produced by the human microbiome play key roles in host-microbe and microbe-microbe interactions and might greatly affect the health state of an individual (1–4). These compounds are often encoded in biosynthetic gene clusters (BGCs) harbored by multitudes of bacterial species populating the human body. The identification and thorough analysis of BGCs provide insights into the

Received: August 15, 2023. **Revised:** October 18, 2023. **Editorial Decision:** October 28, 2023. **Accepted:** October 30, 2023

biosynthetic capabilities of the microbiome and ultimately lead to the discovery of pharmaceutically relevant compounds (5–10). Community initiatives such as the Human Microbiome Project (11,12) accumulated vast volumes of sequencing data whose biosynthetic potential still has to be explored.

The genome mining software paved the way for the creation of databases of BGCs computationally predicted from voluminous genomics datasets (13–17). Some of these databases, such as antiSMASH-DB (13), IMG-ABC (14) and BiG-FAM (15), collect BGCs from a wide variety of sources, while others target specific environments, such as the human gut (16) or ocean microbiome (17). However, there is still no dedicated human microbiome BGC database going beyond the most studied human gut environment and spanning multiple body sites. Furthermore, most of the existing databases were created with already outdated software and do not benefit from the latest advances in genome mining techniques.

To address these limitations, we developed ABC-HuMi, the Atlas of BGCs in the Human Microbiome. Our resource accumulates data from three major human microbiome sequence databases and represents BGCs originating from five human body sites and systems. We employed the most advanced BGC mining and analysis software to populate the database and created a user-friendly interactive platform for exploring the collected BGCs and their associated functionalities.

Data retrieval and processing

Data sources

ABC-HuMi integrates data obtained from the Genomic Catalogue of Earth's Microbiomes (JGI GEM (18)), EMBL-EBI MGnify catalogs (19), and the Human Microbiome Project (HMP, (20)). The EMBL-EBI MGnify Unified Human Gastrointestinal Genome v2.0.1 and Human Oral v1.0 catalogs contain >290 000 isolates and MAGs from the human microbiome clustered into 4744 and 452 species representatives, respectively. For further processing, we retrieved the representative genomes (https://www.ebi.ac.uk/metagenomics/browse/genomes). Selecting only the representative genomes as the input is a compromise between performance and sensitivity. This allowed us to keep the total number of BGCs small, but our experiments show that we still cover the vast majority of the metabolomic diversity present in the skipped genomes (75% at the gene cluster family (GCF) level and 98% at the gene cluster clan (GCC) level) (Supplementary Figure S1).

JGI GEM contains >52 000 MAGs from a wide range of environmental and host-associated microbiomes. We retrieved all human microbiome-associated MAGs, excluding the ones associated with the human gut, already covered by the EBI data (https://portal.nersc.gov/GEM/genomes/).

HMP contains metagenomes from more than 3808 samples associated with different parts of the human body. From this dataset, we selected all metagenomes associated with respiratory, skin, and urogenital samples (https://portal.hmpdacc.org/). We also included 1692 HMP reference genomes available at NCBI (BioProject PRJNA28331) ranging in quality from complete genomes to draft contig assemblies. The genomes without associated body site metadata or related to the diseased patients were not considered.

Processing pipeline

As an initial processing step, we harmonized metadata obtained from the source databases and unified the body site categories. For EMBL-EBI MGnify catalogs, the body site was retrieved from the metadata of associated NCBI BioSamples. Metagenome assemblies were taxonomically annotated using Kraken2 (v2.1.2; command-line arguments (cla): −−confidence 0.1) and its PlusPF database (v20220908) (21).

The retrieved genomes were processed with antiSMASH (v7.0.0; cla: −−genefinding-tool=prodigal-m −−asf) (22). A BGC was classified as fragmented if it is located on a contig edge and complete otherwise. All BGCs with the same body site and taxonomy were grouped and MMseqs2 (v14.7) (23) was used to compare BGCs inside each group. We then merged BGCs with a minimum sequence identity of 0.95 into a single database entry with combined metadata. Finally, BGCs were clustered with BiG-SCAPE (v1.1.5; cla: −−mibig) (24) using Pfam (v35.0) (25). We predicted antibiotic resistant genes located within all BGCs at three reliability levels (Perfect/Strict/Loose) using the Resistance Gene Identifier (RGI) software (v6.0.3; cla: −−include_loose) of the Comprehensive Antibiotic Resistance Database (CARD) database (v3.2.8) (26). GC content of BGCs and the corresponding full genome sequences was computed with BioPython (v1.5.3) (27).

Overview of the collected data

The resulting database comprises 19 218 BGCs grouped into 8989 GCFs and 294 GCCs. Figure 1 shows the composition of the database with regard to the associated body site and product types of the BGCs.

Web server implementation

For the implementation of the ABC-HuMi web server, we set up a Django Python web framework (https://djangoproject.com/) and a PostgreSQL database (https://www.postgresql.org/) in docker containers (https://www.docker.com/) with the help of a Cookiecutter template (https://cookiecutter.readthedocs.io/). For the search job query, we are using the task queue manager Celery (http://docs.celeryproject.org) together with the in-memory data structure store Redis (https://redis.io/). The search jobs are handled with BLAST+ (28) using the Biopython wrapper (https://biopython.org/docs/dev/api/Bio.Blast.Applications.html) and cblaster (29) using a Snakemake pipeline (30). The database browsing table is using DataTables (https://datatables.net/) and the Cytoscape network visualization (31) is using cytoscape.js (https://js.cytoscape.org/). The front end of the website also uses Bootstraps (https://getbootstrap.com/) and Font Awesome (https://fontawesome.com/) for design purposes and jQuery (https://jquery.com/) as a utility library.

Database functionality

Versatile search options

ABC-HuMi provides the BLAST+ (28) search for identifying homologs of user-provided sequences in the database, for example, antibiotic resistance genes (Figure 2A, B). The function handles both nucleotide and protein sequences and can also search translated nucleotide sequences of the database BGCs using a protein query (the tBLASTn mode). Since the single gene search is rarely informative for identifying homologous

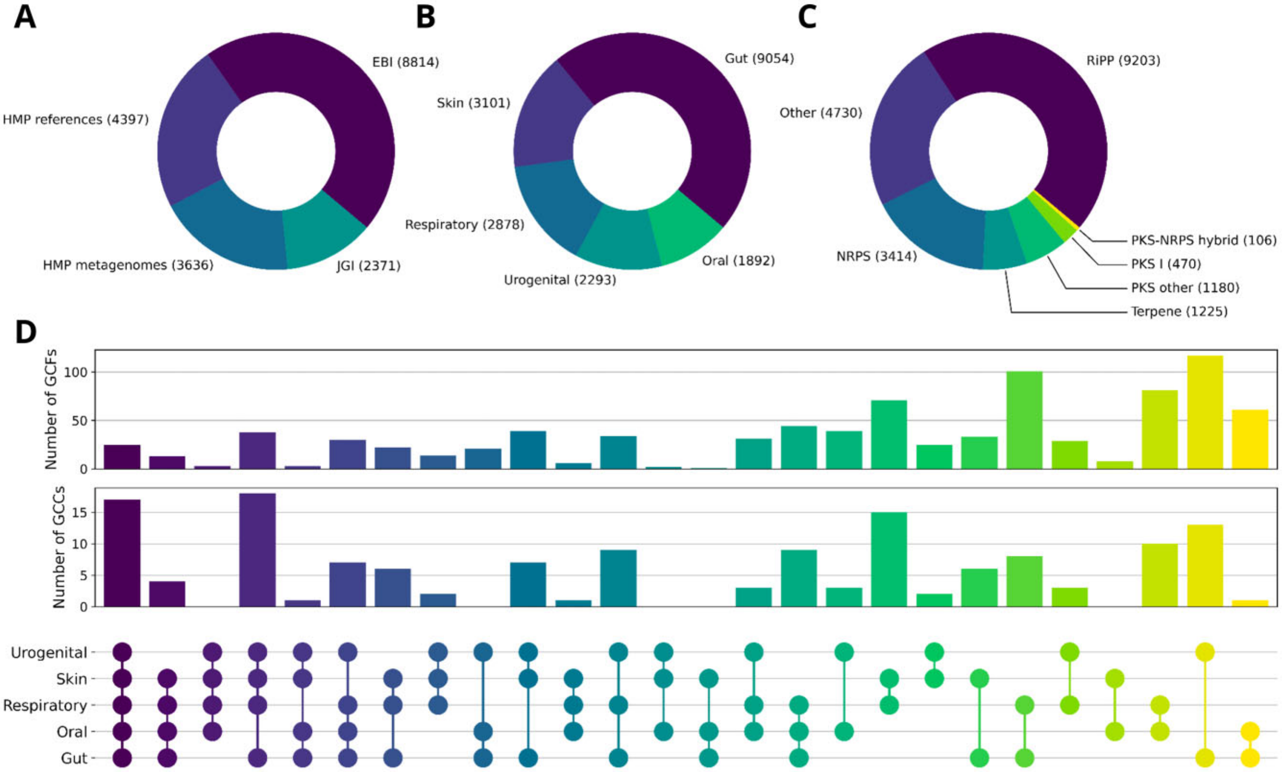


Figure 1. Overview of the ABC-HuMi content. The distributions of biosynthetic gene clusters (BGCs) by (**A**) data source, (**B**) human body site, and (**C**) product type. (**D**) The number of gene cluster families (GCFs) and clans (GCCs) that span across at least two different body sites.

gene clusters (32), we employ cblaster (29) for screening a custom BGC against the entire database (Supplementary Figures S2 and S3).

Interactive tables and visualizations

All BGCs in ABC-HuMi are enriched with detailed metadata that augments antiSMASH-predicted properties with the source human microbiome sequence database and body site, the taxonomy of the producing organism, and the links to similar BGCs within our database and in MIBiG (33). The information is stored in interactive tables that allow filtering by metadata and seamless switching between BGCs and related gene cluster families (GCFs) and clans (GCCs) (Figure 2B, D). Each BGC is complemented with the fully-embedded antiSMASH report; GCF and GCC networks are visualized with Cytoscape (Figure 2C, E).

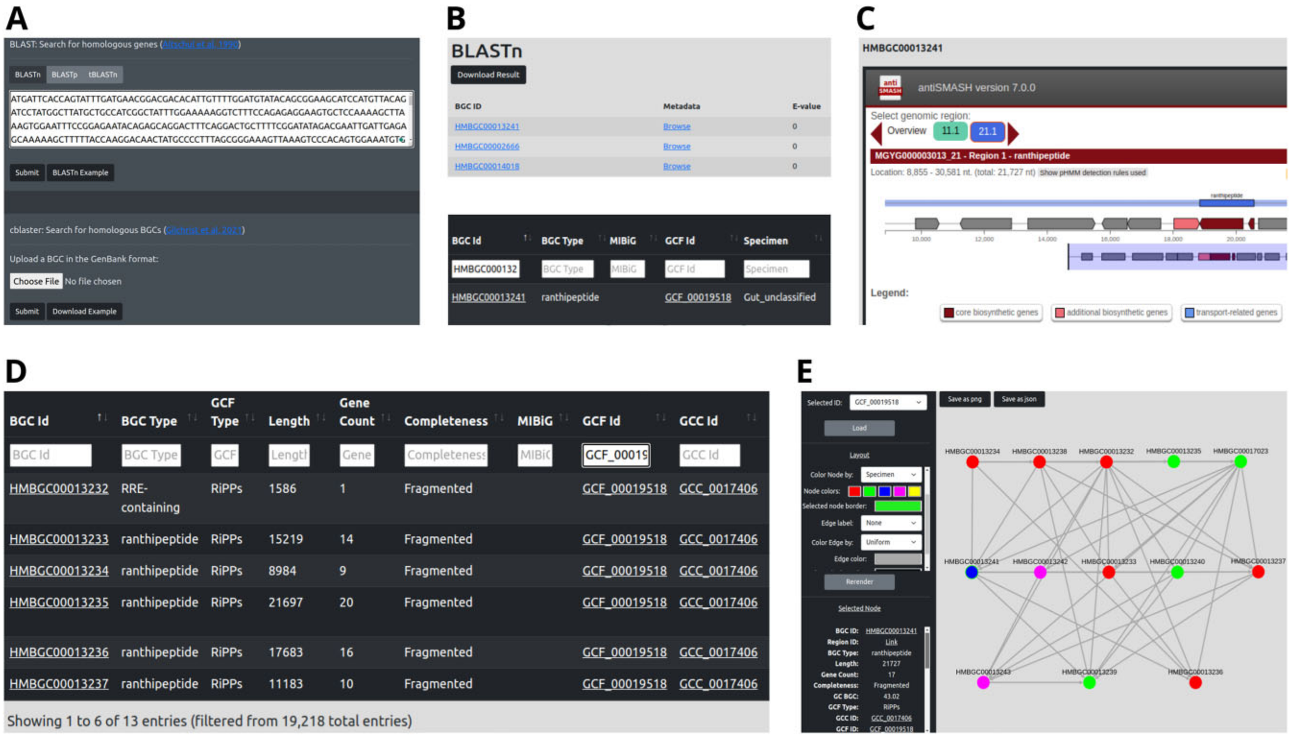
Application examples

Exploring individual BGCs

To explore the hidden bioactive potential of GCFs spanning across multiple human body sites, we employed the versatile ABC-HuMi search functionality. First, we used metadata filtration and Cytoscape visualizations to identify GCFs containing experimentally-validated MIBiG BGCs whose direct neighbors originated from distinct body sites. We further selected two out of 11 hits since the corresponding MIBiG BGCs are known for producing compounds with antimicrobial properties: bacteriocins gassericin T (BGC0000619) and gassericin E (BGC0001388). Both compounds were originally discovered in *Lactobacillus gasseri* (34), one of the main *Lactobacillus* species identified in the vaginal, gastrointestinal, and oral microbiomes (35). The cblaster search of BGC0000619 and BGC0001388 revealed four BGCs in ABC-HuMi with identifiers HMBGC00006519-22 (Supplementary Figures S2–S3). The body sites associated with these BGCs are gut, oral and urogenital system, thus perfectly matching the common localization of the known gassericin producer. Notably, three of the ABC-HuMi BGCs were predicted from *L. paragasseri*, never reported as a gassericin E producer before. Given the resemblance between these two bacterial species, such production is plausible and the ABC-HuMi BGCs might harbor a new variant of the bacteriocin. Though, this computational hypothesis requires experimental validation.

Large-scale comparison

To demonstrate the ABC-HuMi applicability to large-scale analysis, we explored BGCs in a recently published dataset of gut microbiota samples derived from patients with Parkinson's disease (36). First, we applied our data processing pipeline to the binned and unbinned metagenomic data from (36) split into three cohorts of individuals according to the study metadata. BGCs identified with antiSMASH in each group were further used as queries for bulk comparisons against the entire ABC-HuMi database (Supplementary Figure S4). The high ratio of unmatched BGCs at the GCF level (32–43%) strikes the enormous diversity of the human gut microbiome which is not fully covered by the MGnify catalogs of representative genomes. The small difference in distributions of matched BGCs in healthy and diseased cohorts agrees with the original study results, where no significant difference in the microbiome diversity of cohorts was discovered (36). At the same time, the ABC-HuMi-



BLAST: Search for homologous genes (Altschul et al. 1990)

ATGATTCACCAGTATTTGATGAACGGACGACACATTGTTTTGGATGTATACAGCGGAAGCATCCATGTTACAG
ATCCTATGGCTTATGCTGCCATCGGCTATTTGGAAAAAGGTCTTTCCAGAGAGGAAGTGCTCCAAAAGCTTA
AAGTGGAATTTCCGGAGAATACAGAGCAGGACTTTCAGGACTGCTTTTCGGATATAGACGAATTGATTGAGA
GCAAAAAGCTTTTTACCAAGGACAACTATGCCCCTTTAGCGGGAAAGTTAAAGTCCCACAGTGGAAATGTG

cblaster: Search for homologous BGCs (Gilchrist et al. 2021)

Upload a BGC in the GenBank format:

No file chosen

BLASTn

BGC ID	Metadata	E-value
HMBGC00013241	Browse	0
HMBGC00002666	Browse	0
HMBGC00014018	Browse	0

BGC Id	BGC Type	MIBiG	GCF Id	Specimen
HMBGC000132	BGC Type	MIBiG	GCF Id	Specimen
HMBGC00013241	ranthipeptide		GCF_00019518	Gut_unclassified

HMBGC00013241

antiSMASH version 7.0.0

Select genomic region:

MGYG000003013_21 - Region 1 - ranthipeptide

Location: 8,855 - 30,581 nt. (total: 21,727 nt)

Legend:

🟥 core biosynthetic genes 🟥 additional biosynthetic genes 🟦 transport-related genes

BGC Id	BGC Type	GCF Type	Length	Gene Count	Completeness	MIBiG	GCF Id	GCC Id
BGC Id	BGC Type	GCF	Length	Gene	Completeness	MIBiG	GCF_00019	GCC Id
HMBGC00013232	RRE-containing	RiPPs	1586	1	Fragmented		GCF_00019518	GCC_0017406
HMBGC00013233	ranthipeptide	RiPPs	15219	14	Fragmented		GCF_00019518	GCC_0017406
HMBGC00013234	ranthipeptide	RiPPs	8984	9	Fragmented		GCF_00019518	GCC_0017406
HMBGC00013235	ranthipeptide	RiPPs	21697	20	Fragmented		GCF_00019518	GCC_0017406
HMBGC00013236	ranthipeptide	RiPPs	17683	16	Fragmented		GCF_00019518	GCC_0017406
HMBGC00013237	ranthipeptide	RiPPs	11183	10	Fragmented		GCF_00019518	GCC_0017406

Showing 1 to 6 of 13 entries (filtered from 19,218 total entries)

Selected ID: GCF_00019518

Color Node by: Specimen

Node colors: 🟥 🟩 🟦 🟪 🟨

Selected node border: 🟩

Edge label: None

Color Edge by: Uniform

Edge color: ⬜

BGC ID:	HMBGC00013241
Region ID:	Link
BGC Type:	ranthipeptide
Length:	21727
Gene Count:	17
Completeness:	Fragmented
GC BGC:	43.02
GCF Type:	RiPPs
GCC ID:	GCC_0017406
GCF ID:	GCF_00019518

HMBGC00013234 HMBGC00013238 HMBGC00013232 HMBGC00013235 HMBGC00017023

HMBGC00013241 HMBGC00013242 HMBGC00013233 HMBGC00013240 HMBGC00013237

HMBGC00013243 HMBGC00013239 HMBGC00013236

Figure 2. ABC-HuMi search, browse, and visualization functionality. The BLAST search (**A**) identifies biosynthetic gene clusters (BGCs) harboring homologs of the user-provided sequence. The search result page (**B**, top) enables browsing of the BGC metadata (**B**, bottom) and the corresponding embedded antiSMASH reports (**C**). The metadata table links to the gene cluster family (GCF) or clan (GCC) of a BGC, thus allowing one to browse all similar clusters (**D**). GCC and GCF networks can also be visualized via interactive Cytoscape plots (**E**). Here, all nodes of a GCF are titled with the BGC identifiers and colored according to the corresponding human body site: the selected blue node is from the gut, green and pink nodes are from the oral microbiome and red ones are from the respiratory system.

based analysis highlights BGCs shared by multiple body sites (11–14%). This additional layer of information was hidden in the original study focused exclusively on the human gut microbiome but might represent interest for further exploration.

Comparison to existing resources

With over 2500 entries, MIBiG is by far the leading community resource for collecting experimentally-verified BGCs (33,37). However, such BGCs are scarce. A more complete—though less accurate—view on the biosynthetic potential of environmental microbiomes relies on genome mining (5), which naturally leads to the emergence of resources storing computationally-predicted BGCs, such as our database. These resources visibly differ in data focus, processing techniques, and functionality provided to the end users. We compare the ABC-HuMi content and features to sBGC-hm, the only human microbiome-specific BGC database to date, and three state-of-the-art general-purpose *in silico* BGC databases (Table 1).

A direct comparison of the total number of BGCs across the databases could be misleading because of the drastic differences in the focus of the resources, the sourced microbiomes, and the types of underlying genome sequences. AntiSMASH-DB includes only BGCs predicted from isolate genomes, other databases also consider MAGs and ABC-HuMi further relies on unbinned metagenome assemblies as the mining source. The latter increases the number of less reliable and fragmented BGC predictions but enables deeper exploration of the hidden potential of the microbiome. ABC-HuMi keeps the information about the genome type behind each predicted BGC, so users can easily adjust the data selection to their research demands.

While all five databases rely on antiSMASH for mining BGCs, the software version varies from v5 (BiG-FAM, antiSMASH-DB and IMG-ABC) to v6 (sBGC-hm) to the latest v7 (ABC-HuMi), which might greatly affect the accuracy and completeness of the predictions (22,38,39). ABC-HuMi, sBGC-hm and BiG-FAM group identified BGCs into gene cluster families (GCFs) and clans (GCCs). The first two resources utilize BiG-SCAPE (24) while the enormous size of BiG-FAM requires the use of more resource-efficient but less accurate BiG-SLICE (40). The resulting BGCs could be downloaded in bulk only from the ABC-HuMi and BiG-FAM websites, but the further use of the BiG-FAM data might be complicated due to the non-standard data format.

All considered databases provide comprehensive filtering by metadata but the search for a custom sequence or BGC is available only in three of them (ABC-HuMi, BiG-FAM, and sBGC-hm) and only ABC-HuMi supports all search types. Though, antiSMASH-DB and IMG-ABC offer advanced data lookups based on a user-defined set of Pfam domains from a predefined list. AntiSMASH-DB can also be indirectly queried for a custom BGC via the ClusterBlast algorithm embedded into the antiSMASH pipeline (41). The visualization facilities vary greatly among the databases but the most informative detailed view of each BGC (the antiSMASH report) is provided only in ABC-HuMi and antiSMASH-DB.

Table 1. Comparison of *in silico* BGC databases. *# HM-related BGCs (GCFs)* stands for the total number of BGCs (GCFs) related to the human microbiome

Category	ABC-HuMi	sBGC-hm	BiG-FAM	antiSMASH-DB	IMG-ABC
General information					
Data focus	Human microbiome	Human gut microbiome	Microbes in general	Microbes in general	Microbes in general
Genome types	isolates, MAGs, metagenomes	isolates, MAGs	isolates, MAGs	isolates	isolates, MAGs
Latest update	2023 (this work)	2023 (16)	2020 (15)	2020 (13)	2019 (14)
Data processing and availability					
BGC calling	antiSMASH v7 (22)	antiSMASH v6 (38)	antiSMASH v5 (39)	antiSMASH v5 (39)	antiSMASH v5 (39)
BGC clustering	BiG-SCAPE (24)	BiG-SCAPE (24)	BiG-SLICE (40)	✗	✗
# HM-related BGCs (GCFs)	14,821 (7346)	36,583 (8004)	>4791 (>811)[a]	unknown	unknown
Bulk data download	✓ (GenBank)	✗	✓ (custom)	✗	✗
Features					
Metadata search	✓	✓	✓	✓	✓
Sequence search	✓ (BLAST+ (28))	✓ (BLAST+ (28))	✗	✗	✗
Cluster search	✓ (cblaster (29))	✗	✓ (BiG-SLICE (40))	✓ (ClusterBlast (41)[b])	✗
BGC detailed view	✓ (antiSMASH)	✗	✓ (custom[c])	✓ (antiSMASH)	✓ (antiSMASH[d])
Extra features	Cytoscape visualization of GCF/GCC networks with export to PNG and JSON	Matrices of gene co-occurrence in the HMP data Visualizations of differences in core gene coverage across HMP samples	Word cloud of Pfam features for each BGC	Cluster search based on NPRS/PKS module queries Taxonomy tree-based browsing	Pfam-based search with ClusterScout (42) Search by chemical attributes (for experimentally verified products)

[a] The exact numbers are known for human gut MAGs only.

[b] Indirectly available via the antiSMASH pipeline and web interface.

[c] BGC annotation into genes and Pfam domains.

[d] Unavailable for part of the database BGCs.

Conclusion

The unprecedented biosynthetic potential of the human microbiome remains underexplored. In response, we unveil ABC-HuMi, a repository of computationally predicted biosynthetic gene clusters (BGCs) from major human body sites and systems. Despite its compact size, the database represents a huge biosynthetic diversity of the human microbiome and provides a convenient framework for its exploration. By looking at BGCs whose GC content is substantially different from the average genome GC, users might detect possible horizontal BGC transfer events. By inspecting antibiotic resistance genes co-located with BGCs, researchers might identify the producers of promising bioactive compounds. By exploring the distribution of similar BGCs across body sites, one can shed light on the general-purpose or niche-specific function of the underlying BGCs.

The streamlined nature of ABC-HuMi facilitates easy and fast analysis queries, even for computationally demanding applications. The database empowers users with the ability to search not only for custom nucleotide and protein sequences but also entire BGCs and enhances the results with interactive browsing and visualization functionality. Unlike most existing BGC databases, the entire ABC-HuMi can be downloaded locally and used for bulk comparison and prioritization of novel BGCs in large-scale studies. We thus believe the created resource will be beneficial for the diverse needs of the rapidly growing community. We plan to maintain and further expand ABC-HuMi in both functionality and coverage of the human microbiome biosynthetic diversity.

Data availability

The code used to collect and process the data can be found at https://github.com/gurevichlab/abc_humi and https://zenodo.org/records/10027902. The resulting annotated BGC sequences in the GenBank format and all metadata are available from the database website. The resource is freely accessible at https://www.ccb.uni-saarland.de/abc_humi/.

Supplementary Data

Supplementary Data are available at NAR Online.

Acknowledgements

The graphical abstract was created with *BioRender.com*. We thank our colleagues at HIPS for testing the ABC-HuMi database functionality and suggesting new features.

Funding

Saarland University [the NextAID project to support the position of A.T.]; computational resources used within this study were financed through the DFG [466168626]. Funding for open access charge: Saarland University.

Conflict of interest statement

None declared.

References

1. Hou,K., Wu,Z.-X., Chen,X.-Y., Wang,J.-Q., Zhang,D., Xiao,C., Zhu,D., Koya,J.B., Wei,L., Li,J., *et al.* (2022) Microbiota in health and diseases. *Signal Transd. Targ. Ther.*, **7**, 135.
2. Donia,M.S., Cimermancic,P., Schulze,C.J., Brown,L.C.W., Martin,J., Mitreva,M., Clardy,J., Linington,R.G. and Fischbach,M.A. (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell*, **158**, 1402–1414.
3. Heilbronner,S., Krismer,B., tz Oesterhelt,H. and Peschel,A. (2021) The microbiome-shaping roles of bacteriocins. *Nat. Rev. Microbiol.*, **19**, 726–739.
4. Rebuffat,S. (2022) Ribosomally synthesized peptides, foreground players in microbial interactions: recent developments and unanswered questions. *Nat. Prod. Rep.*, **39**, 273–310.
5. Medema,M.H., de Rond,T. and Moore,B.S. (2021) Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.*, **22**, 553–571.
6. Nakatsuji,T., Chen,T.H., Narala,S., Chun,K.A., Two,A.M., Yun,T., Shafiq,F., Kotol,P.F., Bouslimani,A., Melnik,A.V., *et al.* (2017) Antimicrobials from human skin commensal bacteria protect against Staphylococcus aureus and are deficient in atopic dermatitis. *Sci. Transl. Med.*, **9**, eaah4680.
7. Kim,S.G., Becattini,S., Moody,T.U., Shliaha,P.V., Littmann,E.R., Seok,R., Gjonbalaj,M., Eaton,V., Fontana,E., Amoretti,L., *et al.* (2019) Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. *Nature*, **572**, 665–669.
8. Bitschar,K., Sauer,B., Focken,J., Dehmer,H., Moos,S., Konnerth,M., Schilling,N.A., Grond,S., Kalbacher,H., Kurschus,F.C., *et al.* (2019) Lugdunin amplifies innate immune responses in the skin in synergy with host- and microbiota-derived factors. *Nat. Commun.*, **10**, 2730.
9. Sassone-Corsi,M., Nuccio,S.P., Liu,H., Hernandez,D., Vu,C.T., Takahashi,A.A., Edwards,R.A. and Raffatellu,M. (2016) Microcins mediate competition among Enterobacteriaceae in the inflamed gut. *Nature*, **540**, 280–283.
10. Heilbronner,S. and Foster,T.J. (2021) *Staphylococcus lugdunensis*: a skin commensal with invasive pathogenic potential. *Clin Microbiol. Rev.*, **34**, e00205-20.
11. Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, **486**, 207–214.
12. Human Microbiome Project Consortium (2012) A framework for human microbiome research. *Nature*, **486**, 215–21.
13. Blin,K., Shaw,S., Kautsar,S.A., Medema,M.H. and Weber,T. (2021) The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res.*, **49**, D639–D643.
14. Palaniappan,K., Chen,I.-M.A., Chu,K., Ratner,A., Seshadri,R., Kyrpides,N.C., Ivanova,N.N. and Mouncey,N.J. (2020) IMG-ABC v. 5.0: an update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. *Nucleic Acids Res.*, **48**, D422–D430.
15. Kautsar,S.A., Blin,K., Shaw,S., Weber,T. and Medema,M.H. (2021) BiG-FAM: the biosynthetic gene cluster families database. *Nucleic Acids Res.*, **49**, D490–D497.
16. Zou,H., Sun,T., Jin,B. and Wang,S. (2023) sBGC-hm: an atlas of secondary metabolite biosynthetic gene clusters from the human gut microbiome. *Bioinformatics*, **39**, btad131.
17. Paoli,L., Ruscheweyh,H.J., Forneris,C.C., Hubrich,F., Kautsar,S., Bhushan,A., Lotti,A., Clayssen,Q., Salazar,G., Milanese,A., *et al.* (2022) Biosynthetic potential of the global ocean microbiome. *Nature*, **607**, 111–118.
18. Nayfach,S., Roux,S., Seshadri,R., Udwary,D., Varghese,N., Schulz,F., Wu,D., Paez-Espino,D., Chen,I.-M., Huntemann,M., *et al.* (2021) A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.*, **39**, 499–509.
19. Richardson,L., Allen,B., Baldi,G., Beracochea,M., Bileschi,M.L., Burdett,T., Burgin,J., Caballero-Pérez,J., Cochrane,G., Colwell,L.J., *et al.* (2023) MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res.*, **51**, D753–D759.
20. Integrative HMP (iHMP) Research Network Consortium (2019) The integrative human microbiome project. *Nature*, **569**, 641–648.
21. Wood,D.E., Lu,J. and Langmead,B. (2019) Improved metagenomic analysis with Kraken 2. *Genome Biol.*, **20**, 257.
22. Blin,K., Shaw,S., Augustijn,H.E., Reitz,Z.L., Biermann,F., Alanjary,M., Fetter,A., Terlouw,B.R., Metcalf,W.W., Helfrich,E.J., *et al.* (2023) antiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.*, **51**, W46–W50.
23. Steinegger,M. and Söding,J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.*, **35**, 1026–1028.
24. Navarro-Muñoz,J.C., Selem-Mojica,N., Mullowney,M.W., Kautsar,S.A., Tryon,J.H., Parkinson,E.I., De Los Santos,E.L., Yeong,M., Cruz-Morales,P., Abubucker,S., *et al.* (2020) A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.*, **16**, 60–68.
25. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L., Tosatto,S.C., Paladin,L., Raj,S., Richardson,L.J., *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
26. McArthur,A.G., Waglechner,N., Nizam,F., Yan,A., Azad,M.A., Baylay,A.J., Bhullar,K., Canova,M.J., De Pascale,G., Ejim,L., *et al.* (2013) The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.*, **57**, 3348–3357.
27. Cock,P. J.A., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B., *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
28. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
29. Gilchrist,C.L., Booth,T.J., van Wersch,B., van Grieken,L., Medema,M.H. and Chooi,Y.-H. (2021) Cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. *Bioinform. Adv.*, **1**, vbab016.
30. Köster,J. and Rahmann,S. (2012) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, **28**, 2520–2522.
31. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
32. Medema,M.H., Takano,E. and Breitling,R. (2013) Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.*, **30**, 1218–1223.
33. Terlouw,B.R., Blin,K., Navarro-Munoz,J.C., Avalon,N.E., Chevrette,M.G., Egbert,S., Lee,S., Meijer,D., Recchia,M.J., Reitz,Z.L., *et al.* (2023) MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.*, **51**, D603–D610.
34. Kasuga,G., Tanaka,M., Harada,Y., Nagashima,H., Yamato,T., Wakimoto,A., Arakawa,K., Kawai,Y., Kok,J. and Masuda,T. (2019) Homologous expression and characterization of gassericin T and gassericin S, a novel class IIb bacteriocin produced by Lactobacillus gasseri LA327. *Appl. Environ. Microbiol.*, **85**, e02815-18.
35. Selle,K. and Klaenhammer,T.R. (2013) Genomic and phenotypic evidence for probiotic influences of *Lactobacillus gasseri* on human health. *FEMS Microbiol. Rev.*, **37**, 915–935.
36. Becker,A., Schmartz,G.P., Gröger,L., Grammes,N., Galata,V., Philippeit,H., Weiland,J., Ludwig,N., Meese,E., Tierling,S., *et al.* (2022) Effects of resistant starch on symptoms, fecal markers, and gut microbiota in Parkinson's disease—the RESISTA-PD trial. *Genomics, Proteomics & Bioinformatics*, **20**, 274–287.
37. Medema,M.H., Kottmann,R., Yilmaz,P., Cummings,M., Biggins,J.B., Blin,K., De Bruijn,I., Chooi,Y.H., Claesen,J.,

Coates,R.C., *et al.* (2015) Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.*, **11**, 625–631.

38. Blin,K., Shaw,S., Kloosterman,A.M., Charlop-Powers,Z., Van Wezel,G.P., Medema,M.H. and Weber,T. (2021) antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.*, **49**, W29–W35.
39. Blin,K., Shaw,S., Steinke,K., Villebro,R., Ziemert,N., Lee,S.Y., Medema,M.H. and Weber,T. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.*, **47**, W81–W87.
40. Kautsar,S.A., van der Hooft,J.J., de Ridder,D. and Medema,M.H. (2021) BiG-SLiCE: a highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *Gigascience*, **10**, giaa154.
41. Medema,M.H., Blin,K., Cimermancic,P., de Jager,V., Zakrzewski,P., Fischbach,M.A., Weber,T., Takano,E. and Breitling,R. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*, **39**, W339–W346.
42. Hadjithomas,M., Chen,I.M.A., Chu,K., Huang,J., Ratner,A., Palaniappan,K., Andersen,E., Markowitz,V., Kyrpides,N.C. and Ivanova,N.N. (2017) IMG-ABC: new features for bacterial secondary metabolism analysis and targeted biosynthetic gene cluster discovery in thousands of microbial genomes. *Nucleic Acids Res.*, **45**, D560–D565.

Received: August 15, 2023. **Revised:** October 18, 2023. **Editorial Decision:** October 28, 2023. **Accepted:** October 30, 2023