



BioMGCore: a toolkit for detection of biological metabolites in microbiome

Guang Yang¹, Guohao Yu¹, Qian Cui¹,
Xueping Huang¹, Kexin Guo¹, Shulei Jia^{2*}

¹ Jiangsu Ocean University, Jiangsu, 222005, China

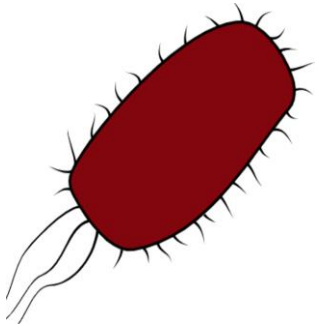
² Tianjin Medical University, Tianjin, 300070, China



Guang Yang, Guohao Yu, Qian Cui, Xueping Huang, Kexin Guo, Shulei Jia. 2025. BioMGCore: a toolkit for detection of biological metabolites in microbiome. *iMetaOmics* 2: e70036. <https://doi.org/10.1002/imo2.70036>



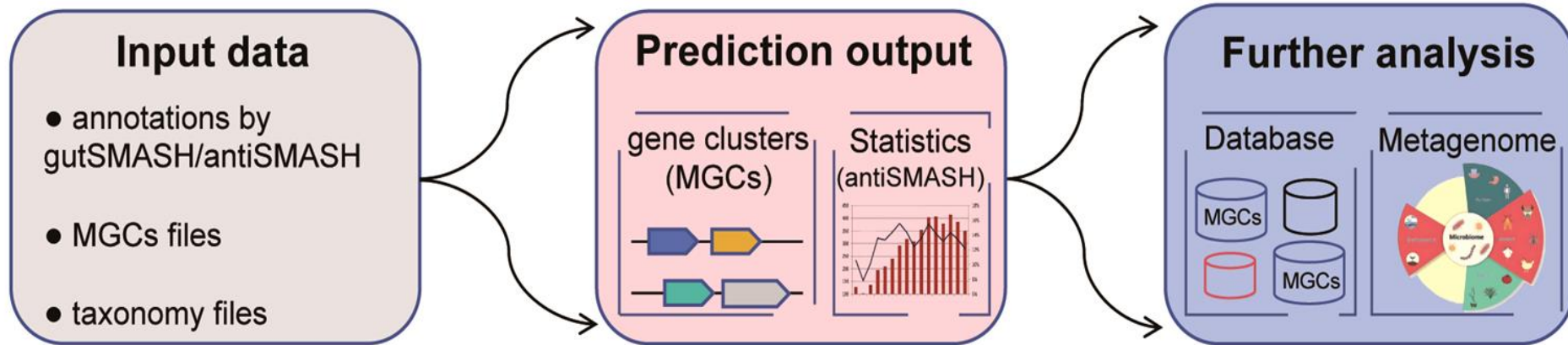
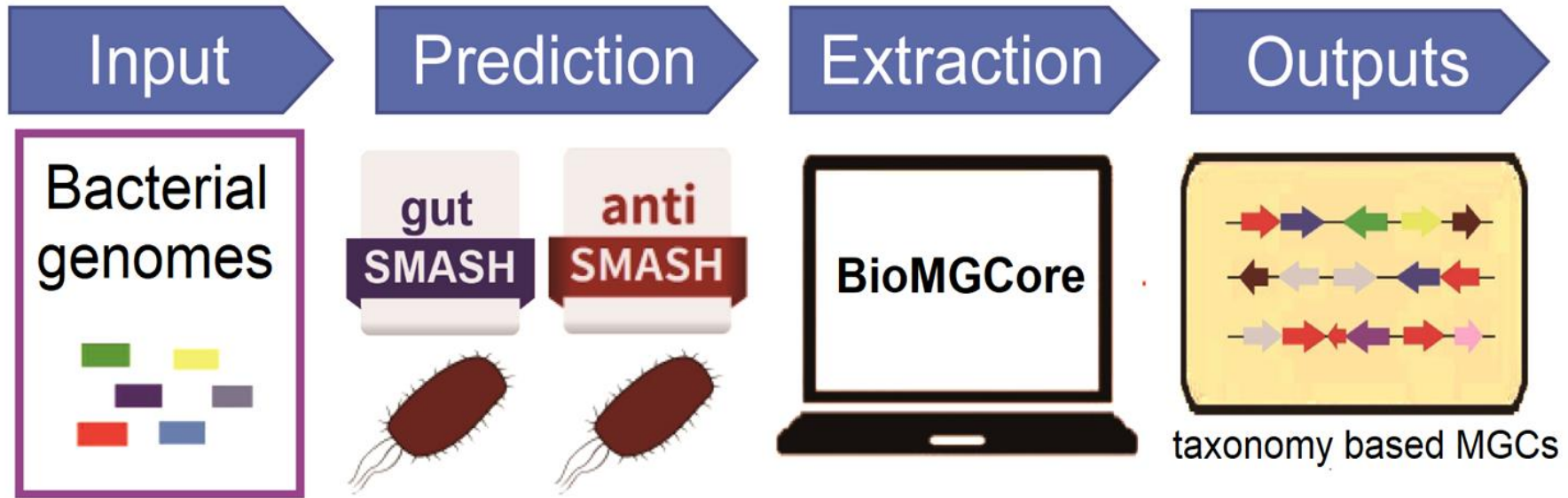
Introduction



- Microbial metabolites are closely related to human health
- Microbial metabolites are encoded by core gene clusters (MGCs), which are tightly arranged and work together
- Microbial primary/secondary metabolite analysis tools are widely used with large and complex annotations
- To solve the problem of processing massive results, the gene cluster extraction tool BioMGCore has been well developed
- The developed toolkit can help explore potential primary or secondary metabolites, greatly improving work efficiency



Highlights



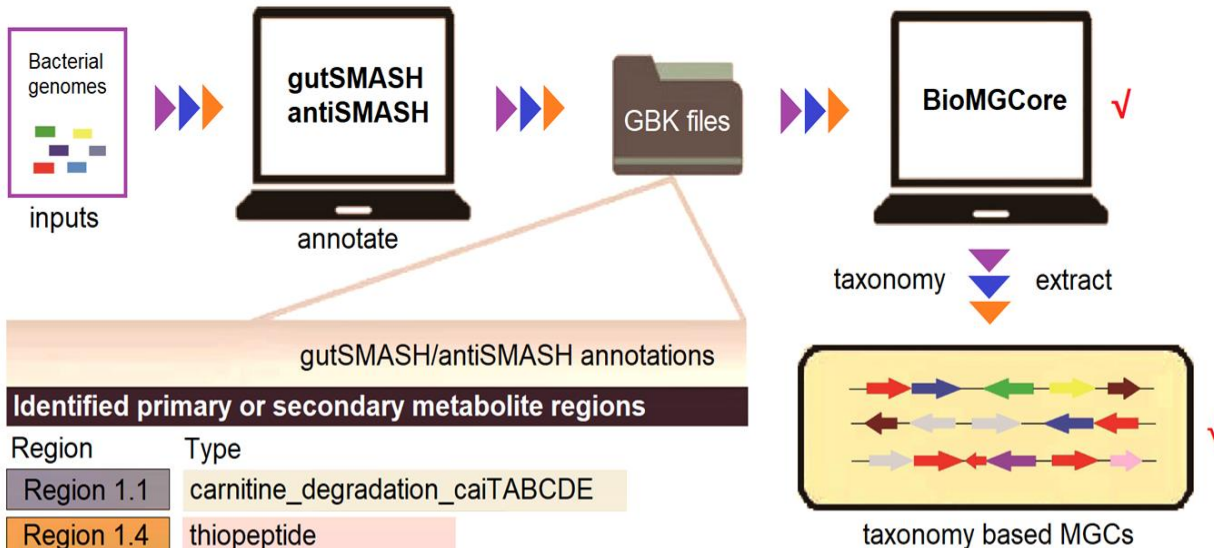
<https://github.com/xielisos567/BioMGCore>



Overview of the workflow

Mainly two steps: annotation of microbial primary/secondary metabolites and running the BioMGCore toolkit.

(A)



(B)

```
usage: BioMGCore.py [-h] -i INPATH -n TYPE -o OUT [-g GAP] [-t {d,p,c,o,f,g,s}]

Get gene info

options:
  -h, --help            show this help message and exit
  -i, --inpath INPATH  input path, original data path
  -n, --type TYPE       gene name
  -o, --out OUT         output dir
  -g, --gap GAP         gap between two genes
  -t, --tax_level {d,p,c,o,f,g,s}
                        taxonomy level (d:Domain, p:Phylum, c:Class, o:Order, f:Family, g:Genus, s:Species)
```

Batch extract core MGCs from annotations:

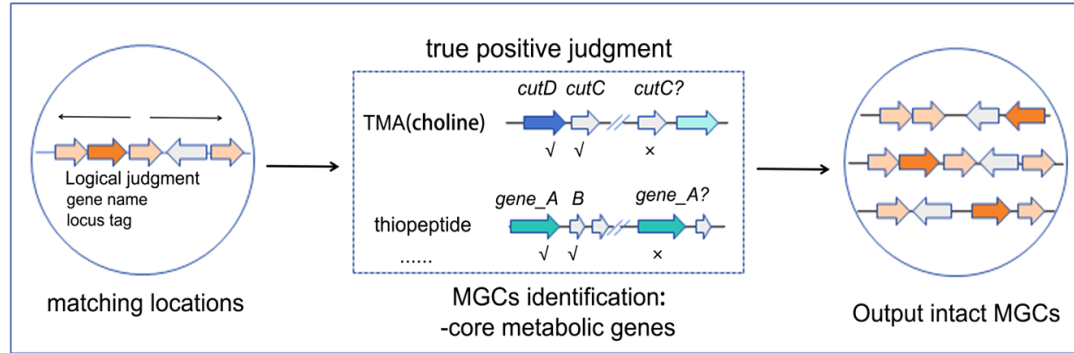
- (1) Accurately identify and extract designated functional gene clusters
- (2) Add source information for taxonomic groups
- (3) Batch statistics for predicting secondary metabolites
- (4) Simultaneous extraction of nucleotides/amino acids

<https://github.com/xielisos567/BioMGCore>



Running the BioMGCore toolkit

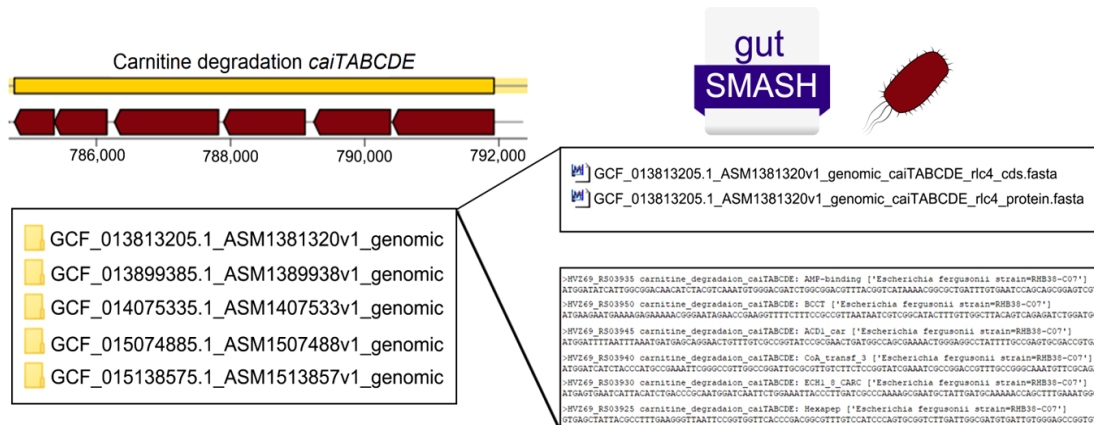
(A)



BioMGCore toolkit running logic:

- (1) Keyword matching logic
- (2) The principle of proximity of genes in MGCs

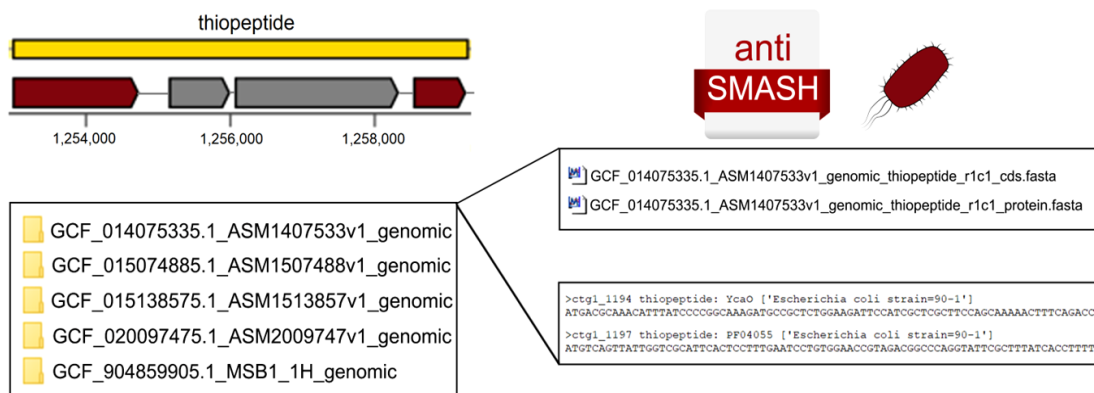
(B)



BioMGCore toolkit extraction:

- (1) Extract according to the gutSMASH annotations
- (2) Extract based on antiSMASH annotations
- (3) Add taxonomic information

(C)





Case 1: Microbial analysis workflows

Table S1. Statistical analysis of antiSMASH annotations with the BioMGCore toolkit

SGBs	Chromosome	Region	Type	From	To	Most similar known cluster	Similarity
AM09-9_scaffold	AM09-9.Scaf5	Region5.1	ranthipeptide	75239	96880		
AM09-9_scaffold	AM09-9.Scaf7	Region7.1	RRE-containing	82492	100828		
AM09-9_scaffold	AM09-9.Scaf23	Region23.1	RRE-containing	1138	21404		
AM09-9_scaffold	AM09-9.Scaf25	Region25.1	RRE-containing	32859	43393		
AM09-9_scaffold	AM09-9.Scaf30	Region30.1	RRE-containing	6752	27018		
AM34-8_scaffold	AM34-8.Scaf7	Region7.1	thiopeptide	126915	153207	O-antigen	14%
AM34-8_scaffold	AM34-8.Scaf14	Region14.1	NRP-metallophore,NRP	1	40003	enterobactin	100%
AM34-8_scaffold	AM34-8.Scaf34	Region34.1	NRPS,T1PKS	1	23002	yersiniabactin	10%
AF27-11AA_scaffold	AF27-11AA.Scaf13	Region13.1	ranthipeptide	13672	35554		
AF29-16BH_scaffold	AF29-16BH.Scaf1	Region1.1	phosphonate	581279	595658	polysaccharide B	6%
AF29-16BH_scaffold	AF29-16BH.Scaf10	Region10.1	ranthipeptide	83986	105216		
TF06-40_scaffold	TF06-40.Scaf15	Region15.1	RRE-containing	63185	83451		
OM05-15BH_scaffold	OM05-15BH.Scaf8	Region8.1	RRE-containing	287188	307454		
OM05-15BH_scaffold	OM05-15BH.Scaf25	Region25.1	RRE-containing	23301	43651		
OM05-15BH_scaffold	OM05-15BH.Scaf28	Region28.1	arylpolylene	10621	47431	flexirubin	48%
AF22-12AC_scaffold	AF22-12AC.Scaf1	Region1.1	RRE-containing	61572	81829		
AF22-12AC_scaffold	AF22-12AC.Scaf8	Region8.1	ranthipeptide	4714	26084		
AF22-12AC_scaffold	AF22-12AC.Scaf1	Region1.2	RRE-containing	84450	105592		
TM07-10_scaffold	TM07-10.Scaf2	Region2.1	RiPP-like	85698	96483		
TM07-10_scaffold	TM07-10.Scaf23	Region23.1	NRPS-like	1	30191		
TM07-10_scaffold	TM07-10.Scaf2	Region2.2	opine-like-metallophore	194688	211727		

Running code:

```
python3 antiSTAT.py -i data_path_antiSMASH_annotation -o ./statistics.xlsx
```

(1) At least 1266 (/1520) bacteria had the potentials to synthesize secondary metabolites, with intact MGCs being detected in their genomes.

(2) Functional MGCs for synthesizing cyclic lactone autoinducers have the highest proportion (31.22%) among all MGCs and coexist with other secondary metabolite genes.

Case 2: Microbial analysis workflows

Table S2. Analysis of gutMASH annotations with the BioMGCore toolkit

Accessions	Species	Strain info
GCF_002208985.1	Citrobacter koseri	strain=FDAARGOS_287
GCF_002393245.1	Citrobacter koseri	strain=FDAARGOS_393
GCF_002947035.1	Citrobacter koseri	strain=AR_0025
GCF_002947675.1	Citrobacter koseri	strain=AR_0024
GCF_003812405.1	Citrobacter koseri	strain=FDAARGOS_530
GCF_008693945.1	Citrobacter koseri	strain=FDAARGOS_646
GCF_014679775.1	Citrobacter koseri	strain=MPUCK001
GCF_016128275.1	Citrobacter koseri	strain=FDAARGOS_1029
GCF_019443705.1	Citrobacter koseri	strain=NJ
GCF_000783755.2	Citrobacter freundii	strain=FDAARGOS_73
GCF_000783995.2	Citrobacter freundii	strain=FDAARGOS_61
GCF_001022155.1	Citrobacter freundii	strain=CAV1321
GCF_001022275.1	Citrobacter freundii	strain=CAV1741
GCF_001702455.1	Citrobacter freundii	strain=B38
GCF_001718055.1	Citrobacter freundii	strain=SL151
GCF_002211705.1	Citrobacter freundii	strain=705SK3
GCF_002215385.1	Citrobacter freundii	strain=18-1
GCF_002786865.1	Citrobacter freundii	strain=CRCB-101
GCF_002796505.1	Citrobacter freundii	strain=AR_0022
GCF_003571565.1	Citrobacter freundii	strain=AR_0116
GCF_003665535.1	Citrobacter freundii	strain=UMH19

(1) A total of 3541 (/25,755) bacterial strains carried intact *caiTABDE* gene cluster, mainly belonging to *Escherichia coli* and *Salmonella enterica*, both members of the Firmicutes phylum.

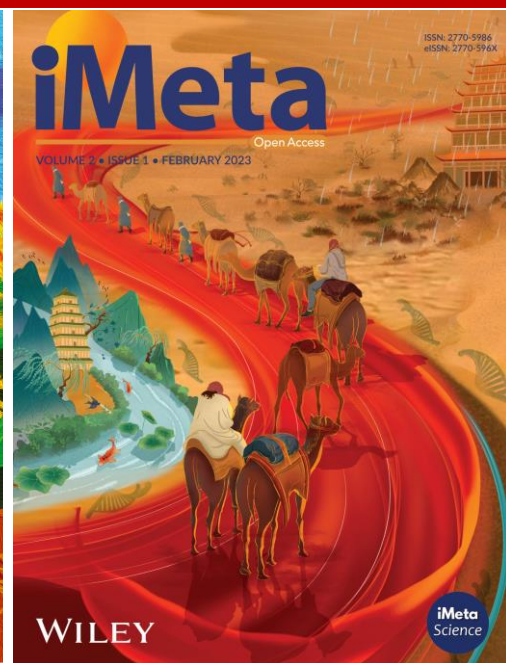
(2) A species-specific database of *caiTABDE* gene cluster was constructed , which was helpful for targeted therapy of CVDs.



Summary

- ❑ In this study, we introduced a user-friendly toolkit designed specifically for the study of microbial primary/secondary metabolites - BioMGCore.
- ❑ At present, BioMGCore can identify and extract core functional gene clusters, and support analysis and statistics based on different taxonomic groups.
- ❑ BioMGCore has been well applied in microbiomic analysis with high accuracy. The tool will continue to be upgraded to meet users' needs for standardized and personalized microbiomic analysis.
- ❑ Website: <https://github.com/xielisos567/BioMGCore>

Guang Yang, Guohao Yu, Qian Cui, Xueping Huang, Kexin Guo, Shulei Jia. 2025. BioMGCore: a toolkit for detection of biological metabolites in microbiome. *iMetaOmics* 2: e70036. <https://doi.org/10.1002/imo2.70036>



“**iMeta**” launched in 2022 by iMeta Science Society, **impact factor (IF) 33.2**, ranking **top 65/22249 in world and 2/161 in the microbiology**. It aims to publish innovative and high-quality papers with broad and diverse audiences. **Its scope is similar to Cell, Nature Biotechnology/Methods/Microbiology/Medicine/Food**. Its unique features include video abstract, bilingual publication, and social media with 600,000 followers. Indexed by **SCIE/ESI, PubMed, Google Scholar** etc.

“**iMetaOmics**” launched in 2024, with a **target IF>10**, and its scope is similar to **Nature Communications, Cell Reports, Microbiome, ISME J, Nucleic Acids Research, Briefings in Bioinformatics**, etc.

“**iMetaMed**” launched in 2025, with a **target IF>15**, similar to **Med, Cell Reports Medicine, eBioMedicine, eClinicalMedicine** etc.



Society: <http://www.imeta.science>

Publisher: <https://wileyonlinelibrary.com/journal/imeta>

iMeta: <https://wiley.atyponrex.com/journal/IMT2>

Submission: iMetaOmics: <https://wiley.atyponrex.com/journal/IMO2>

iMetaMed: <https://wiley.atyponrex.com/journal/IMM3>



[iMetaScience](#)



[iMetaScience](#)



office@imeta.science
imetaomics@imeta.science



[Promotion Video](#)

Update
2025/6/18