



BioMGCore: 用于检测微生物组中生物代谢产物的工具包

杨光¹, 俞国豪¹, 崔仟¹, 黄雪萍¹, 郭可欣¹, 贾书磊^{2*}

¹ 江苏海洋大学

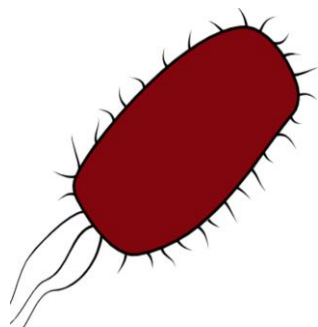
² 天津医科大学



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>



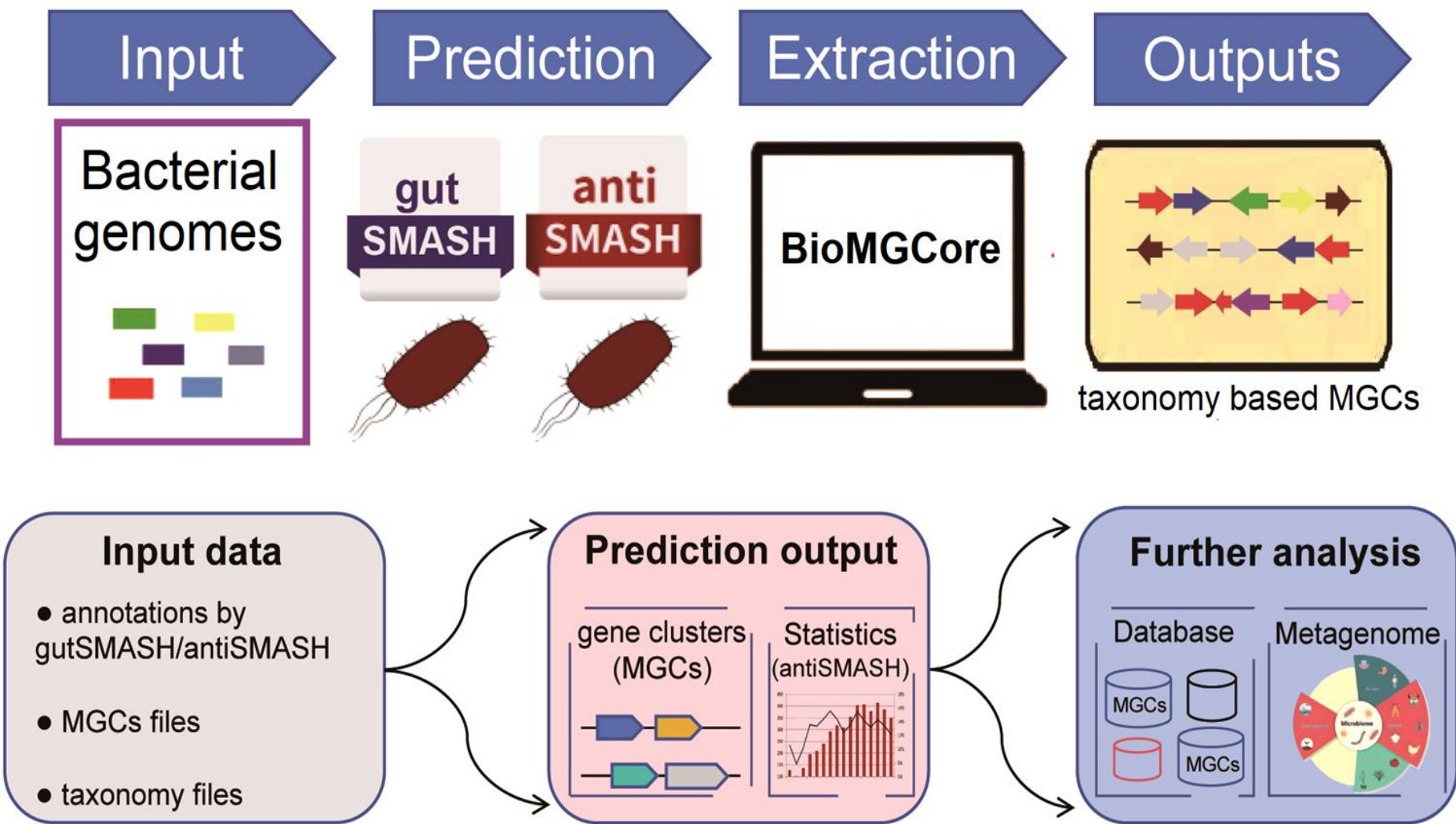
简介



- 微生物代谢产物与人体健康密切相关
- 微生物代谢产物是由核心基因簇（MGCs）编码的，这些基因紧密排列共同发挥作用
- 微生物初级/次级代谢产物分析工具已得到广泛应用，但注释结果庞大而繁杂。
- 为解决海量结果的处理问题，我们开发了基因簇提取工具BioMGCore
- 开发的工具包能够帮助探索潜在的初级或次级代谢物，极大提高了工作效率



亮点



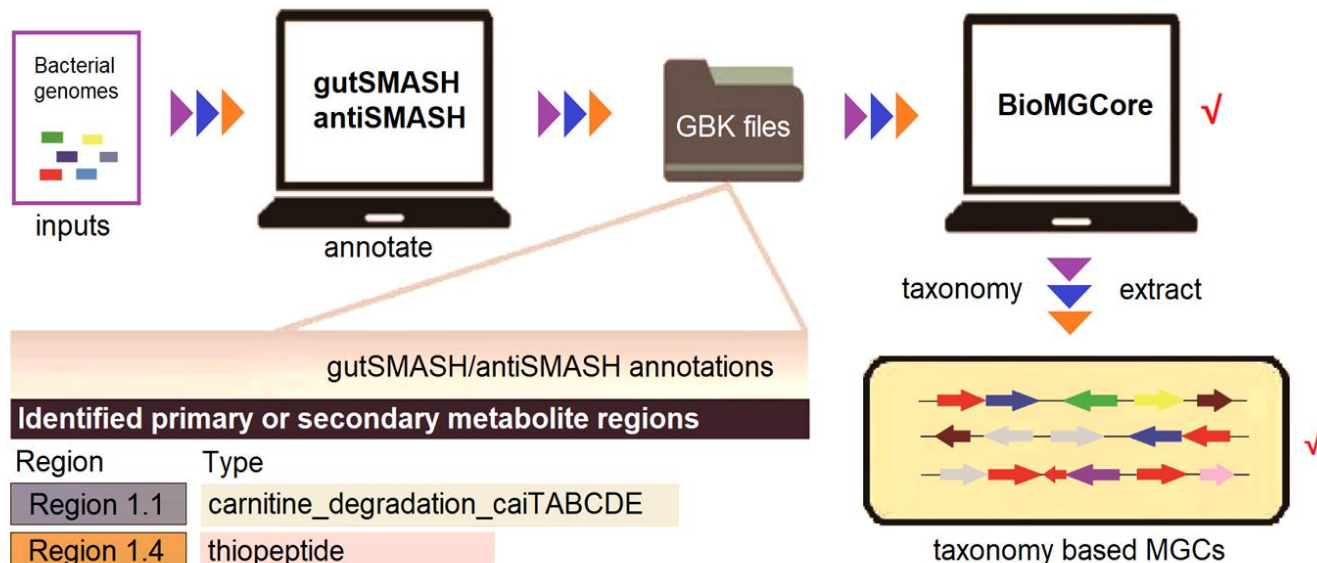
<https://github.com/xielisos567/BioMGCore>



流程概述

主要包含两个步骤：微生物初级/次级代谢物注释和BioMGCore工具包运行。

(A)



从注释结果中批量提取核心基因簇：

- (1) 精准识别并提取指定的功能基因簇
- (2) 可添加taxonomy分类群来源信息
- (3) 实现对预测次级代谢产物的批量统计
- (4) 核苷酸/氨基酸序列同步提取

(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)
```

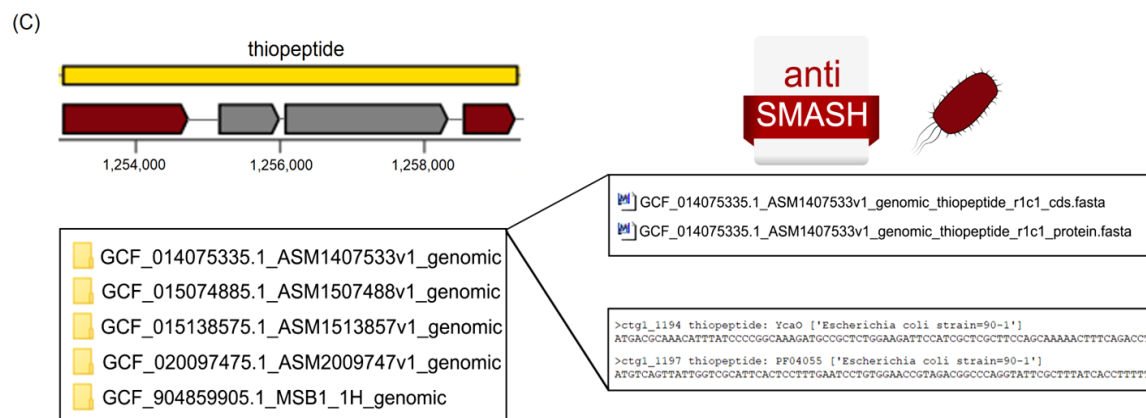
<https://github.com/xielisos567/BioMGCore>



(2) 基因簇中基因的邻近原则



(3) 添加物种分类群信息





案例一：微生物组数据分析流程

Table S1. 使用BioMGCore工具包对antiSMASH注释结果统计

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	arylpolyene	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		

运行代码：

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

(1) 对1520个细菌基因组的分析发现至少有1266个细菌具有合成次生代谢物的能力，在基因组中检测到了完整MGCs。

(2) 用于合成环内酯自诱导物（cyclic-lactone-autoinducer）的功能性MGCs在所有MGCs中所占比例最高（31.22%），且与其他次生代谢物基因共存。



案例二：微生物组数据分析流程

Table S2. 使用BioMGCore工具包对gutSMASH注释结果分析

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）在25,755个参考细菌基因组中对心血管疾病相关的代谢物γ-丁酰甜菜碱(γ-butyrobetaine)分析发现，共有3541株细菌菌株含有完整的*caiTABCDE*基因簇，主要属于*Escherichia coli*和*Salmonella enterica*，二者均为厚壁菌门成员。

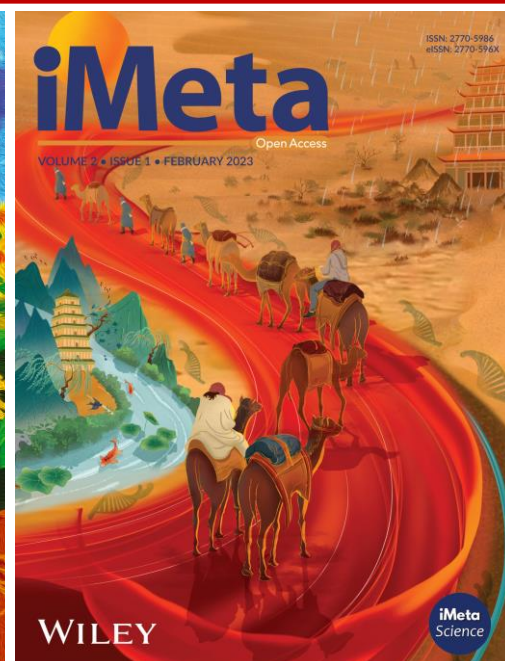
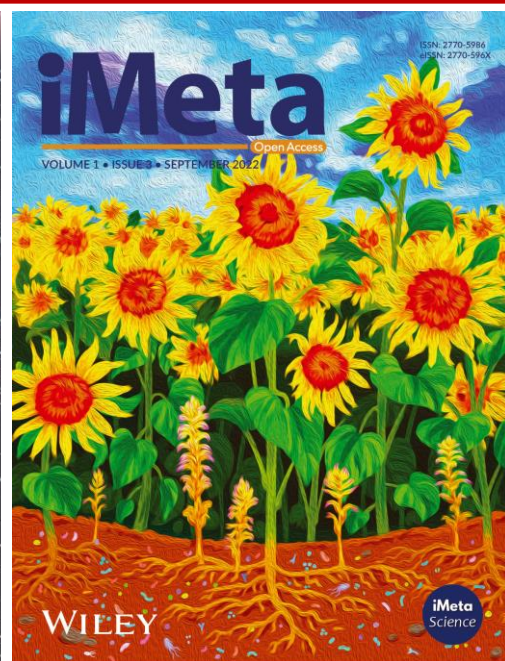
（2）构建了一个*caiTABCDE*基因簇的物种特异性数据库，这有助于心血管疾病的靶向治疗。



总结

- ❑ 在这项研究中，我们介绍了一个专为微生物初级/次级代谢产物研究设计的用户友好型工具包——BioMGCore;
- ❑ 目前BioMGCore能够实现对核心功能基因簇的识别与提取，并支持按不同物种分类群进行分析统计;
- ❑ BioMGCore在微生物组数据分析中应用良好，具有较高的准确性，该工具将持续升级，满足用户开展标准化和个性化的微生物组学数据分析;
- ❑ 工具包网址: <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期刊(影响因子**33.2**)由宏科学、千名华人科学家和威立出版, 主编刘双江和傅静远教授。目标为生物/医学/环境综合期刊群(对标Cell), 高影响力的研究、方法和综述欢迎投稿, 重点关注生物技术、大数据和组学等前沿交叉学科, 已被SCIE、PubMed等收录, IF 33.2位列全球SCI期刊第65(前千分之三), 微生物学研究类全球第一, 中国第5, 中科院生物学双1区Top; 外审平均21天, 投稿至发表中位数87天。

子刊**iMetaOmics**(主编赵方庆和于君教授)、**iMetaMed** 定位IF>10和15的综合、医学期刊, 欢迎投稿!



主页: <http://www.imeta.science>

出版社: <https://wileyonlinelibrary.com/journal/imeta>

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

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

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



office@imeta.science

imetaomics@imeta.science



宣传片



iMeta



更新日期
2025/6/18