



人体肠道噬菌体的研究与应用

刘聪^{Δ1,4} 邢博^{Δ1,2} 陈婉桐^{1,3} 李俊桦^{*1,4} 肖敏凤^{*1,4}

1 深圳华大生命科学研究院 广东 深圳 518083

2 中国科学院大学生命科学学院 北京 100049

3 华南理工大学生物科学与工程学院 广东 广州 510006

4 深圳市未知病原体应急检测重点实验室 广东 深圳 518083

摘要: 肠道是人体内微生物定殖最丰富的部位。近年来,随着肠道菌群与人体健康疾病关联研究的蓬勃发展,肠道噬菌体也逐渐引起关注。然而,相关信息技术和实验技术发展的滞后在一定程度上限制了肠道噬菌体的科学研究进程。因此,本文首先回顾了近几年来肠道噬菌体研究领域所开发或采用的计算和实验方法,包括噬菌体的测序数据分析和噬菌体的分离纯化等。随后,本文就肠道噬菌体的分类、肠道内噬菌体与细菌的互作及肠道噬菌体在人体疾病干预中的应用展开了讨论。最后,本文展望了肠道噬菌体研究在数据和实体资源、信息和实验技术、与肠道菌群的互作、干预和治疗人体疾病各方面的一系列挑战和机遇。

关键词: 病毒组,噬菌体数据库,肠道噬菌体挖掘,噬菌体库,微生物疗法,肠道微生态,慢病干预

Research and application of human gut bacteriophage

LIU Cong^{Δ1,4} XING Bo^{Δ1,2} CHEN Wantong^{1,3} LI Junhua^{*1,4} XIAO Minfeng^{*1,4}

1 BGI-Shenzhen, Shenzhen, Guangdong 518083, China

2 College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

3 School of Biology and Biological Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China

4 Shenzhen Key Laboratory of Unknown Pathogen Identification, Shenzhen, Guangdong 518083, China

Abstract: The gut is the largest reservoir of human flora. In the last decade, with the increasing investigations on the relationships between gut microbiota and human health diseases, gut bacteriophage has also drawn more attention. However, the research of gut bacteriophage is still in its infancy, hampered by delayed progress in the experimental and bioinformatical technology for gut bacteriophage and virome. Thus, this review first summarizes the research methodology of gut bacteriophage in computation and experiments,

Foundation item: National Natural Science Foundation of China (82050007)

ΔThese authors equally contributed to this work

*Corresponding authors: Tel: 86-755-33945504

E-mail: LI Junhua: lijunhua@genomics.cn; XIAO Minfeng: xiaominfeng@genomics.cn

Received: 29-05-2021; **Accepted:** 04-08-2021; **Published online:** 24-08-2021

基金项目: 国家自然科学基金(82050007)

Δ对本文贡献相同

*通信作者: Tel: 0755-33945504

E-mail: 李俊桦: lijunhua@genomics.cn; 肖敏凤: xiaominfeng@genomics.cn

收稿日期: 2021-05-29; 接受日期: 2021-08-04; 网络首发日期: 2021-08-24

including sequencing analysis and phage isolation, etc. Then, phage taxonomy, phage-host interactions, and the clinical applications of gut bacteriophage are also discussed. Finally, this review looks forward to the challenges and opportunities for gut bacteriophage research in data and physical resources, bioinformatical and experimental technology, interactions with intestinal bacteria, interventions on human diseases, etc.

Keywords: virome, phage database, gut phage mining, phagebank, microbial therapy, gut microecology, chronic disease intervention

人体内含有 10^{14} 个微生物, 与人体细胞数量相当, 而人体内微生物定殖最丰富的部位是肠道^[1]。肠道微生物包括细菌、病毒、真菌、古生菌等, 它们在肠道中参与人体的代谢、免疫等生理过程, 对人体的健康起着不可或缺的作用^[2], 被誉为一个独立的器官^[1]。2014 年, Li 等建立了当时最全的肠道微生物参考基因组^[3]。肠道菌群的组成、丰度和多样性等特征及其扰动和变化与人体的健康与疾病的状态如肥胖、2 型糖尿病、心血管疾病、肠癌、类风湿性关节炎等存在强相关性, 通过口服益生菌或抑菌药物干预肠道菌群可以有效地改善疾病状态^[4-10]。与此同时, 人体肠道细菌的病毒, 即噬菌体也引起了人们的关注。人体每克粪便样本中病毒样颗粒的数量约有 10^9-10^{10} 个, 其中主要成分是噬菌体^[11]。研究表明, 它们可以通过侵染宿主菌和水平转移基因等方式改变人体肠道菌群, 进而对人体的健康状态和疾病发生起着至关重要的作用^[12-14]。

随着测序技术的成熟, 人们尝试直接对样本中所有病毒进行测序, 进而探索人体肠道病毒组^[15]。许多病毒组的数据信息已存储在公共数据库中, 后续研究者通过与这些数据库进行比对, 可对样本中病毒组的数据信息有初步了解。然而目前已知肠道病毒序列有限, 在与公共数据库比对后可被鉴定到的肠道噬菌体序列多是被称为“暗物质”的未知病毒^[11]。同时, 由于病毒基因组缺乏通用标记基因, 只依赖数据库比对的方法会将研究范围局限在可识别的小部分病毒序列中。因此, 通过测序数据分析方法可以挖掘更多新颖的病毒序列^[16-17]。然而, 即便挖掘更多肠道噬菌体基因组序列可以补充肠道噬菌体数据资源, 噬菌体实体资源的缺乏则限制

了噬菌体基因组学与其生物学特征关联分析的进展, 制约了肠道噬菌体与宿主菌甚至与人体免疫系统相互作用的研究发展, 影响了肠道噬菌体对人体慢性疾病的干预探索等。因此人体肠道噬菌体颗粒的获取尤为重要。

通常, 分离肠道噬菌体颗粒首先需要获得其宿主菌, 而肠道菌体外分离培养的实验操作条件较为苛刻。随着微生物培养组学技术的发展, 挑食性(Fastidious)细菌逐渐揭开神秘的面纱, 我国深圳华大生命科学研究院、英国 Wellcome Trust Sanger 研究所、中国科学院微生物研究所这 3 家机构均构建了大型肠道菌库^[18-20], 这为分离获取肠道噬菌体提供了宝贵的资源。此外, 部分肠道噬菌体特有的生物特征严重制约了肠道噬菌体颗粒的获取^[21-24], 如 crAssphage 在分离过程中无明显噬菌斑出现。尽管目前已有研究者分离纯化出数十株肠道噬菌体, 并研究其侵染宿主菌的机制。然而肠道噬菌体数量庞大, 欲探究肠道噬菌体群落及其与肠道微环境、人体健康和疾病的相互关系, 我们仍需持续完善肠道噬菌体组学数据资源和实体资源^[22]。

总而言之, 新技术催生新格局。宏基因组学和培养组学技术给人体肠道菌群的研究与应用带来了变革。同样地, 技术发展也有助于肠道噬菌体的研究与应用实现突破。2015 年, 高通量测序技术和组学信息技术在人体肠道病毒组的研究领域初试锋芒; 2017 年起, 每年均有新的噬菌体生物信息学工具问世, 推动了肠道噬菌体研究的高速发展, 包括发现了全新的噬菌体和噬菌体亚科, 揭示了噬菌体与肠道菌的相互作用, 挖掘了噬菌体组与人体疾病的关联, 尝试了针对慢性疾病的噬菌体疗

法等。下文将针对人体肠道噬菌体的研究方法, 人体肠道噬菌体的实体资源、数据资源及病毒学分类, 肠道噬菌体与宿主菌互作, 肠道噬菌体在慢性疾病干预中的应用这四大部分展开综述(图 1)。

1 人体肠道噬菌体的研究方法

1.1 人体肠道噬菌体测序信息分析方法

随着测序技术的进步, 包括噬菌体核酸在内的宏病毒组学也迅速发展。宏病毒组学是宏基因组学的一个分支, 结合病毒自身特点, 将病毒与其他微生物分开, 提取样本中所有病毒的核酸构建宏病毒组文库, 利用宏基因组测序的方法研究样本中病毒组的遗传信息。2003 年, Breitbart 等使用鸟枪测序的方法对人类粪便中未培养的病毒群落进行宏病毒组学分析发现, 可识别的病毒主要是噬菌体, 整个肠道病毒组约包含 1 200 种病毒基因型^[15]。随后, 大规模人体肠道病毒组的研究加深了人们对人体肠道病毒组组成及多样性的认知。人体肠道噬菌体的组成具有高度异质性, Shkorporov 等证实了人体肠道病毒高度的个体特异性和稳定性^[25]; Gregory 等对健康西方人处于生命周期中不同阶段的肠道

病毒组进行分析发现婴儿和成年人肠道病毒组总丰度较高, 而老年人肠道病毒组丰度显著减少^[26]; Ma 等利用生物信息学方法证实了人体肠道噬菌体组和 2 型糖尿病的关联性^[27]。此外, 饮食、地理位置、健康状态等因素均会对人体肠道病毒组产生很大的影响^[28-30]。最新的研究结果显示, 新生儿的肠道病毒组呈逐步定殖的规律, 随着时间的推移, 肠道病毒组多样性和丰度显著增加, 该规律与肠道菌群在新生儿体内的定殖模式相似^[31-32]。尽管对人体肠道病毒组的研究日益增加, 但仍有高达 90% 的人体肠道病毒序列与当前参考数据库几乎无同源性。Clooney 等分析了炎症性肠病患者肠道内已知和未知病毒序列的组成和功能变化, 提出关于健康和疾病人中肠道病毒暗物质的观点^[33]。Camarillo-Guerrero 等利用 28 060 个全球分布的人体肠道宏基因组和 2 898 个培养的肠道细菌参考基因组, 共发现超过 14 万种病毒基因组, 并利用其进行了人体肠道噬菌体的多样性、细菌宿主预测和全球分布等分析^[34]。Benler 等基于已公开的人体肠道宏基因组数据, 通过搜索携

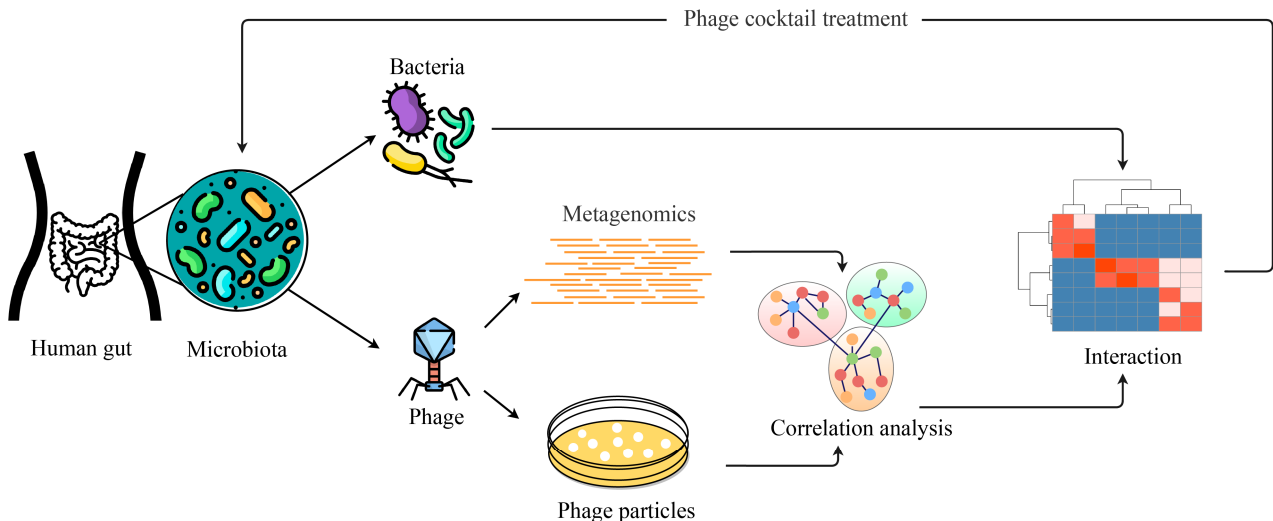


图 1 肠道噬菌体资源挖掘、遗传多样性、生物学特征流程及应用

Figure 1 The resource mining, genetic diversity, biological characteristics and applications of human gut bacteriophage

Note: Phage particles reveal phage resource mining; metagenomics reveals phage prediction and the diversity of genome; combine phage particles with phage genome information show the phage biological characteristics; the research on phage and host interaction is to explore the applications of gut phage and manipulate the human microbiota

带编码噬菌体标志性基因的环形序列, 鉴定出 3 738 个类似完整的噬菌体基因组, 其中有 3 个噬菌体属与已知噬菌体属差别较大, 可能成为新的噬菌体属^[35]。

1.1.1 测序数据质控及组装

宏病毒组研究对人体肠道噬菌体, 特别是未培养的人体肠道噬菌体的研究至关重要。但是高通量测序(Massively Parallel Sequencing, MPS)产生的大量数据给质控和序列分析带来极大挑战。因此, 一些应用于宏基因组/宏病毒组的工具和软件应运而生(表 1)。宏病毒组生物信息学分析一般分为预处理、过滤、组装、病毒序列鉴定和后续分析 5 个步骤(图 2)。2018 年的一篇综述对 49 种应用于宏病毒组学分析的生物信息学工具进行了讨论, 其中 SRSA、Exhaustive Iterative Assembly、VIP 这 3 个工具整合了所有分析步骤^[57]。宏病毒组测序数据

下机后, 对其进行基础质控过滤, 过滤后对获得的 Clean Data 的测序质量进行评估并对数据产出进行统计。值得注意的是, 虽然可以采用病毒颗粒富集技术, 但是并不能完全避免宿主序列污染, 因此还需对宿主序列信息剔除, 如对于人类粪便样本, 可通过和人类完整参考基因组比对去除人基因组污染。去污染后的序列通过组装可获得更完整的病毒序列, 便可对病毒基因进行深度解读。肠道病毒组的组装主要依赖于短序列从头组装(*de novo* Assembly), 但常会导致组装碎片化或病毒群落成员恢复不完全^[58]。2019 年的一项研究对迄今为止用于病毒组组装的软件进行了全面的测评, metaSPAdes 和 MEGAHIT 在病毒组组装上都展示出较好的性能, metaSPAdes 组装的准确性更好、组装片段更长; MEGAHIT 使用内存较小、运行速度较快, 可根据样本数量和计算内存对 2 个软件进

表 1 常用宏病毒组生物信息学分析工具

Table 1 The list of common bioinformatic analysis tools

工具 Tools	功能 Functions	参考文献 References
FastQC	Quality control	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Fastp	Quality control	[36]
Trimmomatic	Quality control	[37]
SOAPnuke	Quality control	[38]
Prinseq	Quality control	http://prinseq.sourceforge.net/index.html
metaSPAdes	Assembly	[39]
MEGAHIT	Assembly	[40]
CD-HIT	Sequence clustering	[41]
MetaPhlan2	Species identification	[42]
Kraken+Bracken	Species identification	[43]
Virusorter2	Virus sequence prediction	[16,44]
VirFinder	Virus sequence prediction	[17]
VirMiner	Virus sequence prediction	[45]
VirusSeeker	Virus sequence prediction	[46]
Metavir 2	Virus sequence prediction	[47]
Prophage Hunter	Prophage prediction	[48]
PHASTER	Prophage prediction	[49]
CheckV	Completeness assessment	[50]
vConTACT	Classification	[51-52]
CRISPRCasFinder	CRISPR spacers prediction	[53]
CRISPRminer	CRISPR spacers prediction	[54]
MinCED	CRISPR spacers prediction	https://github.com/ctSkennerton/minced
PredPHI	Host prediction	[55]
PHP	Host prediction	[56]

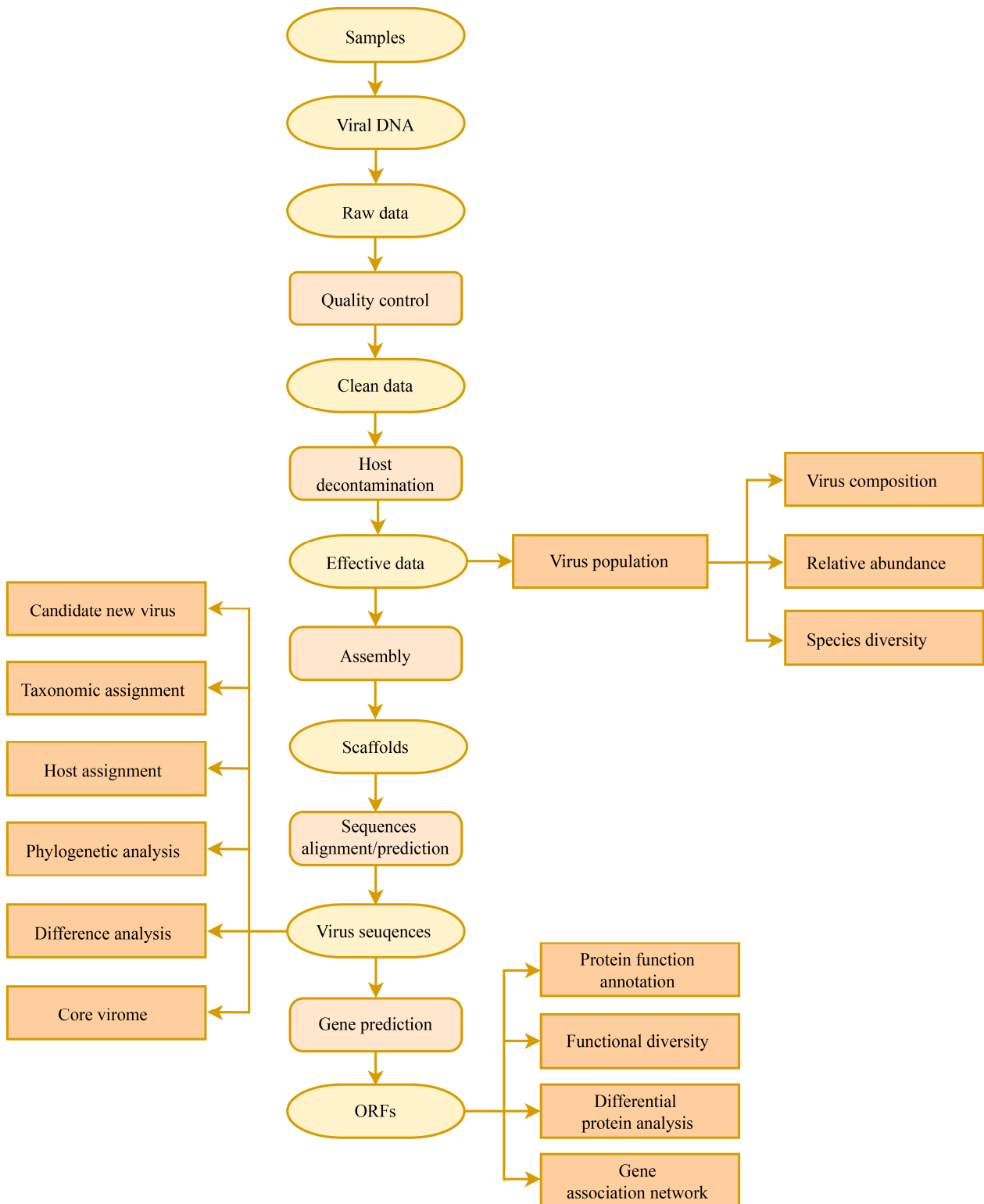


图 2 宏病毒组生物信息学分析流程

Figure 2 The pipeline of bioinformatic analysis of metavirome

行选择^[59]。由于目前应用于病毒组组装的软件各有优劣,同时病毒组在样本中的相对丰度低且多数病毒序列未知,因此开发出全面的、针对病毒组的组装软件尤为重要。

1.1.2 数据比对与噬菌体预测

得到组装的序列后,可通过和已公开病毒基因组数据库进行比对鉴定病毒序列,常用的数据库有 nt 库(Nucleotide Sequence Database, http://www.genefinity.org/sp/sp_nucdatabases.html)、IMG/VR (Integrated Microbial Genome/Virus System, <https://img.jgi.doe.gov/cgi-bin/vr/main.cgi>)、RefSeq (Reference Sequence Database, <https://www.ncbi.nlm.nih.gov/refseq/>)、GVD (Gut Virome Database)、GPD (Gut Phage Database)等^[26,34,60-61]。但仅依赖于与公开数据库进行比对的方法已无法满足快速增加的病毒序列分析需求。因此,多个病毒序列预测的应用工具不断涌现。Virsorter2、VirFinder、VirMiner 等新兴工具均使用不同模型对宏基因组数据中的病毒序列进行预测,其中 Virsorter2 可以对烈性噬菌体和前噬菌体进行区别,并按照预测的可信程度进行分类^[16-17,44-45]。Prophage Hunter 是首个以预测前噬菌体活性为特色的工具,整合了序列相似性比对和机器学习方法,可在一定程度上识别整合于细菌基因组的前噬菌体是否经历了变异事件从而失活^[48]。利用该工具,我们可基于目标肠道菌的全基因组序列预测前噬菌体区间,评估前噬菌体可被诱导的可能性,并对其基因功能进行注释。这给前噬菌体的诱导实验提供了较精准的指导,提高了实验成功率,给积累肠道噬菌体数据和实体资源提供了便利。Nayfach 等开发了对病毒序列进行质量和完整性评估的工具——CheckV,可通过将预测得到的病毒序列与完整的病毒基因组的大型数据库进行比较来评估其完整性,该软件还可用于区分前噬菌体和非前噬菌体,当宿主序列完整性小于10%时,该软件准确性可达到92.8%,远远超过现有的其他软件^[50]。

1.1.3 噬菌体物种鉴定及其宿主预测

通过比对和预测得到的病毒序列可用于后续的分析,包括在多条病毒序列之间、病毒序列与其他物种序列之间以及不同队列中病毒组之间的数据分析。一些常用的宏基因组序列物种鉴定软件也可用于病毒物种的鉴定。MetaPhlan2 软件基于序列相似性,使用 BLAST 或 Bowtie 等比对软件,通过参考物种的 Marker 序列进行比对,达到物种鉴定的目的。Kraken 软件通过计算序列中包含的特殊的 *k*-mer 序列为其分配物种信息,与 MetaPhlan2 相比, Kraken 的速度更快,获取的物种分类也更多。由于病毒序列的占比很少且大多数序列都是未知的,因此上述软件的准确性较低。Bolduc 等开发了一个使用基因共享网络对未培养的病毒序列进行分类的工具——vConTACT,该软件在属层面对病毒序列的分类与国际病毒分类学委员会(The International Committee on Taxonomy of Viruses, ICTV)的分类标准有 96%的重合^[51-52]。CRISPRCasFinder、CRISPRminer 等软件可利用 CRISPR-Cas 系统对细菌 CRISPR 间隔子序列和 Cas 蛋白序列进行预测,将预测到的病毒序列与之进行 BLASTn 比对,可实现对噬菌体宿主的预测^[53-54]。Li 等开发了一种基于深度学习可根据序列数据预测噬菌体宿主的工具——PredPHI^[55]。Lu 等开发了一种原核病毒宿主预测器的软件——PHP,是以病毒与宿主基因组序列之间的 *k*-mer 频率差异为特征对原核病毒的宿主进行预测^[56]。Marbouty 等利用宏基因组 Hi-C 技术将宏基因组组装基因组与噬菌体序列连接起来,对肠道中噬菌体与宿主重叠群进行表征^[62]。此外,Alawi 等开发了可用于队列样本进行病毒组分析的工具——DAMIAN,该软件可以通过比对和预测的方式快速准确地检测病毒序列,并可比较多个样本中共有的病毒序列,可用于传染病暴发时病原微生物快速鉴定^[63]。

1.2 人体肠道噬菌体颗粒获取方法

尽管包括人体肠道噬菌体在内的病毒组学研

究越来越多,但肠道噬菌体颗粒的获取仍然较少,导致肠道噬菌体与宿主菌、与肠道微环境相互作用机制等研究进展缓慢。因此,对人体肠道噬菌体颗粒的需求已迫在眉睫。用于肠道噬菌体分离的样品主要来自粪便。由于微生物组在肠道内不同位置的组成丰度不同,可用小型动物的肠道黏膜和结肠样本,或者其他可能富有靶标菌的环境样品比如污水、海水、土壤等进行噬菌体颗粒挖掘^[64]。其中,对于样本量较大的样品比如污水、海水,可用 0.45 μm 或 0.2 μm 聚醚砜(Polyethersulfone, PES)滤膜过滤除菌后直接加入一定比例的浓缩培养基,接种靶标菌进行噬菌体生物富集^[65]。除此之外,也可以先对样品进行物理浓缩再进行后续试验,即将污水物理浓缩后,与等比的培养基、目标靶向菌进行共同培养,从而达到噬菌体生物富集的目的。样品的物理浓缩主要有超速离心(100 000 $\times g$ 离心 1 h)、使用 10% (质量体积比) PEG8000-1 mol/L NaCl 沉降病毒颗粒重悬浓缩和超滤浓缩(30 kD 或 100 kD 孔径超滤膜)^[22,66]。对样品的物理浓缩在开展噬菌体生物富集前后均可。浓缩后的病毒组无论是进行双层平板法观察噬菌斑,还是提取核酸进行测序分析都更加灵敏。

对于样品量稀少或取样困难的样品,如粪便、组织、肠内容物等来源样本,通常需要浸泡匀浆于 Sodium Chloride-Magnesium Sulfate (SM)缓冲液,离心后的上清用 0.45 μm 或 0.2 μm PES 滤膜过滤

除菌。由于样品液过滤后噬菌体数量可能有所降低,因此通常需要进行目标噬菌体的生物富集,即用处理后的样品与对数期的单株或多株靶标菌混合共培养数小时或过夜,以达到富集目标噬菌体的目的。随后,离心上清过滤除菌的样本分别和单株靶标菌混合富集培养,重复上述单株菌富集步骤 3-5 轮,但上述方法完全依赖于实验室已有菌株的数目和多样性。近年研究者利用生物信息学技术对样本中可能含有的菌种、噬菌体及噬菌体靶向宿主进行预测^[23]。因此,在利用样本进行噬菌体筛选的同时可以对同一来源的样本进行宿主菌株的筛选(图 3)。Guerin 等采用粪便发酵的方法将粪便均质液不经除菌处理直接接种于厌氧发酵罐内,通过添加抗生素,选择性抑制了革兰氏阳性菌和碱性厌氧菌的生长,对样本中原有的拟杆菌门细菌及其靶向噬菌体进行原位富集,随后利用从同一粪便样本中分离得到的细菌进行常规靶向噬菌体富集分离筛选实验^[24]。粪便发酵方法不再受限于实验室保藏菌种库,而是任由粪便中的细菌和噬菌体原位富集生长,有效提高了噬菌体筛选分离的效率,扩大了实验样品体积,更方便于湿实验同时进行,提高实验容错率。

在对采集的样本进行了生物富集或物理浓缩后,需要验证噬菌体的富集效果并将其分离纯化。常规噬菌体的分离纯化使用点滴法或双层琼脂法获取噬菌斑,连续挑取单噬菌斑进行纯化。然而在

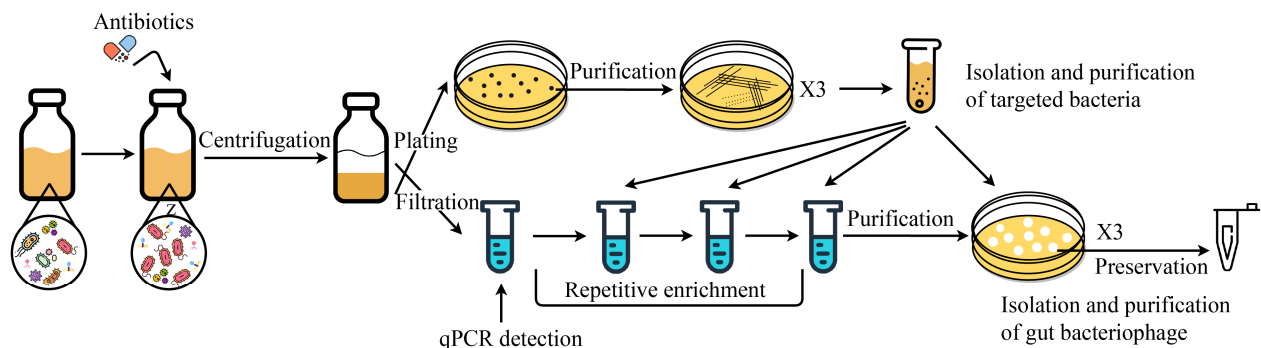


图 3 肠道噬菌体分离纯化流程

Figure 3 The pipeline of isolation and purification of gut bacteriophages

对肠道噬菌体内占比达 90% 的 crAss 样噬菌体的分离实验中发现, 通过噬菌斑观察裂解性不强的噬菌体的富集效果不佳, 温和性噬菌体和不裂解宿主菌的丝状噬菌体也难以用传统的噬菌体分离方法获取颗粒^[23-24]。然而研究者们可对富集液上清进行宏基因组测序或 qPCR 监测其中细菌和噬菌体的浓度变化, 以快速反映噬菌体富集效果, 提高噬菌体获取成功率^[24]。

2 人体肠道噬菌体的分类

人体肠道是病毒定殖最丰富的器官, 其中主要成分是噬菌体。对获得的肠道噬菌体实体和测序得到的肠道噬菌体序列进行分类, 可以更好地了解肠道噬菌体的物种多样性和遗传多样性^[11]。

2.1 基于病毒形态学的肠道噬菌体分类

ICTV 对原核病毒的传统分类是基于病毒形态学进行的。人体肠道噬菌体由遗传物质不同分为单链 DNA (ssDNA) 噬菌体和双链 DNA (dsDNA) 噬菌体, 其中小型 ssDNA 噬菌体主要为微小噬菌体科 (Microviridae) 和丝状噬菌体科 (Inoviridae), dsDNA 噬菌体主要为有尾噬菌体目 (Caudovirales) 的长尾噬菌体科 (Siphoviridae)、短尾噬菌体科 (Podoviridae) 和肌尾噬菌体科 (Myoviridae)^[67]。一项针对 5 000 多人人体肠道宏基因组数据的分析显示, 人体肠道噬菌体中绝大部分(78%)都属于有尾噬菌体目, 有尾噬菌体目根据不同的头部和尾部形态划分为不同的噬菌体科^[35]。然而, 基于形态学的分类

方法其准确性受到质疑。Sutton 等提出, 具有相似基因组和功能特征的噬菌体(如噬菌体 P22 与噬菌体 λ) 依据病毒体形态被分为不同的噬菌体家族^[68]。

2.2 基于基因组特征的肠道噬菌体分类

ICTV 在 2020 年提出了一种新的 15 级病毒分类学分类体系, 对从宏基因组学数据中得到的病毒核酸序列进行分类, 以适应病毒的遗传多样性。基于全基因组平均核苷酸同源性 (Average Nucleotide Identity, ANI) 进行物种等级分类的方式被广泛应用, 将序列覆盖率超过 85% 和 ANI 超过 95% 设为阈值, 对肠道噬菌体组序列聚类成噬菌体簇后, 依据系统发育关系或/和基因共享网络对其进行分类, 进而发现候选的噬菌体新分类^[69] (表 2)。(1) 基于系统发育关系: 噬菌体基因组上含有一些编码标志蛋白的基因, 根据噬菌体标志蛋白构建系统发育树, 可以对噬菌体, 特别是未培养的噬菌体进行分类。Dutilh 等在研究人体肠道微生物时, 偶然发现了一些在人体肠道宏基因组中分布广泛且含量丰富的有尾噬菌体目的双链 DNA 病毒 crAssphage, 它们和已知病毒序列差别较大, 使用保守或广泛存在的噬菌体蛋白构建进化树发现, crAssphage 属于高度独立的一个进化分支, 被证实为一种新型病毒^[70]。Devoto 等从成年人肠道宏基因组数据中鉴定到一类基因组特别大、长度超过 540 kb 的噬菌体 Megaphage, 依据大终止酶基因构建系统发育树表明, Megaphage 属于肌尾噬菌体科, 但在系统发育树上位于单独的进化支, 结合它们巨大的基因组

表 2 基于基因组特征分类的人体肠道噬菌体

Table 2 Human novel intestinal phages based on the classification of genomic characterizations

名称	鉴定方法	基因组大小	宿主	参考文献
Name	Identification methods	Genome size (kb)	Hosts	References
crAssphage	Phylogenetic tree	~97.0	Bacteroides	[70]
Lak phages	Phylogenetic tree	~540.0	Prevotella	[71]
Gubaphage	Phylogenetic tree	~80.0	Bacteroides and Parabacteroides	[34]
Flandersviridae	Phylogenetic tree and gene-sharing networks	~85.2	Bacteroides and Parabacteroides	[35]
Gratiaviridae	Phylogenetic tree and gene-sharing networks	~107.4	Bacteroides and Parabacteroides	[35]
Quimbyviridae	Phylogenetic tree	~75.2	Prevotella, Bacteroides, and Parabacteroides	[35]

和不同的遗传密码, 作者认为其为新的噬菌体分类, 并命名为 Lak Phages^[71]。另一项关于人体肠道宏基因组的研究发现了一组和 crAssphage 具有相似特征的噬菌体序列, 其在人体肠道宏基因组中被鉴定到的数量仅次于 p-crAssphage, 而且在全球范围内广泛分布; 使用大终止酶基因构建系统发育树发现, 这组噬菌体与其他 crAssphage 明显不同, 形成了一个独特的进化支, 作者将这一噬菌体进化支称为肠道拟杆菌噬菌体(Gubaphage)^[34]。(2) 基于基因共享网络: 将相关病毒簇基因组作为一节点, 通过病毒簇之间的相关性进行连接, 建立基因共享网络以达到对未培养噬菌体分类的目的^[52]。Shkoporov 等基于基因共享网络发现占人体肠道噬菌体大多数的是有尾噬菌体目下的长尾噬菌体科、肌尾噬菌体科以及近几年新发现的 crAssphage, 它们在基因共享网络中呈相互关联的分布状态, 而相对富集的微小噬菌体科在网络中形成非常紧密的孤立簇^[25]。(3) 基于系统发育关系和基因共享网络: 近年的一些研究将系统发育关系和基因共享网络结合起来, 更好地应用于肠道噬菌体分类及新噬菌体的发现。最新的一项对大规模人类肠道宏基因组的研究, 通过对编码大终止酶亚基(TerL)、衣壳或门户蛋白等噬菌体标志蛋白的环形序列进行分析, 鉴定出 3 738 个较为完整的肠道噬菌体基因组, 并发现了 3 个候选的新噬菌体科 Flandersviridae、Quimbyviridae 和 Gratiaviridae^[35]。此外, 通过 vConTACT 构建的基因共享网络发现, 3 个候选的新噬菌体科中的 Flandersviridae 和 Gratiaviridae 与其他已知噬菌体簇间的连接很弱, 即这些噬菌体的基因含量与先前已知噬菌体的基因含量差异较大, 表明它们可能是新的噬菌体科^[35]。

3 肠道噬菌体与宿主菌互作

3.1 噬菌体生命周期及其转换

传统上, 噬菌体的生命周期因侵染宿主菌所用的侵染方法不同而通常分为裂解周期(Lytic

Cycle)、溶源周期(Temperate Cycle)和假溶源性(Pseudolysogeny)^[72]。噬菌体与宿主菌之间的相互作用处于一种动态的变化。其中, “杀死胜利者”(Kill-The Winner)模型描述了噬菌体及宿主菌之间的多样性和丰度的快速变化, 即随着优势细菌被其噬菌体杀死, 群落中其他细菌会接替成为优势细菌, 而随后也会被其对应的噬菌体杀死, 以维持群落中噬菌体与宿主菌的多样性和丰度^[73]。除此之外, 在部分微生物群落环境中, 噬菌体裂解周期与溶源周期的转换与微生物群落中噬菌体-细菌比例(Virus-To-Microbe Rate, VMR)有关, 即借助“胜利者”(Piggyback-The-Winner)模型^[74]。该模型提出, 随着细菌丰度的增加, 噬菌体将从烈性向温和性转变, 利用细菌的增殖而达到扩增的目的。一般情况下, 该模型适用于VMR较高的微生物群落中, 但仍有研究表明, 在部分VMR较低的环境中, 噬菌体也出现了借助“胜利者”模型^[75]。虽然人体肠道中的VMR较低, 但该模型在人体肠道中的适用性仍有待进一步研究。

此外, 由于肠道中微生物组的多样性及环境的复杂性, 人体肠道中噬菌体的生命周期也存在更多的可能。如, 人体肠道中最丰富的代表性噬菌体 crAssphage 与温和型噬菌体类似, 不会限制宿主菌的生长和增殖且可与宿主菌以高丰度共存, 然而, 对 crAssphage 基因组分析并没有发现其具有任何与溶源性相关的基因; 同时, 也未在其宿主拟杆菌的基因组中观察到 crAssphage 的前噬菌体序列^[23]。通过对拟杆菌属的噬菌体与宿主菌的动力学研究, 即拟杆菌荚膜多糖(Capsular Polysaccharide, CPS)相变和菌株的瞬时抗性时发现, crAssphage 是以裂解方式复制的, 而 crAssphage 在裂解宿主菌后并无噬菌斑被检测到, 其原因解释如下: (1) 宿主菌的高丰度可能是由瞬时抗性宿主菌维持的, 这些抗性克隆会随着时间的推移而恢复到敏感状态^[22-23]; (2) crAssphage 可能会对宿主菌在菌株或者亚菌株水平发生

Kill-The Winner 的动态变化,但在宿主菌的属水平则可能随着时间的推移保持稳定性,同时可能还参与 Piggyback-The-Winner 的动态模式^[64,74,76]。鉴于此, crAssphage 与拟杆菌的相互作用机制为研究人类肠道中噬菌体与宿主相互作用及其在微生物组中的作用都提供了新的思路,其中宿主种群的表型异质性可能是研究某些烈性噬菌体与其宿主菌在微生物组中稳定互作的核心机制,同时以噬菌斑为鉴定标准判断噬菌体的存在是有局限性的。

3.2 肠道烈性噬菌体与宿主菌的相互作用

除了噬菌体与宿主细菌的比例,噬菌体颗粒在不同空间的分布差异导致其与宿主细菌的接触率不同,从而影响其相互作用,并可能影响烈性噬菌体在肠道中存在的持久性。在一项烈性噬菌体与其宿主细菌共存的小鼠肠道模型实验中,肠黏膜中的烈性噬菌体丰度较低,宿主细菌的计数较高,而在肠腔内则相反;这种空间异质性使得噬菌体与宿主细菌在肠腔和黏膜之间形成了浓度梯度差异,从而限制了烈性噬菌体的侵染和细菌抗性的产生,使烈性噬菌体与其宿主细菌得以共存^[77]。不仅如此,在肠道中宿主菌的可逆相位变化对烈性噬菌体的影响也引起了人们的关注。Jiang 等发现在肠道菌中普遍存在一个称为“倒位子”且能发生可逆性倒位的基因间区域,含有启动子的倒位子大多可调节胞外产物,如荚膜多糖,并且“倒位子”可以使其调控的基因表达状态在“开”和“关”之间转换^[78]。而多形拟杆菌的荚膜多糖的相变可以介导烈性噬菌体对拟杆菌的侵染,即多形拟杆菌可以通过改变相变脂蛋白的表达来逃脱噬菌体的侵染^[22]。其中,肠道菌中“倒位子”的调控是否影响真实肠道中噬菌体与宿主菌的互作仍有待进一步研究。

3.3 肠道温和性噬菌体与宿主菌的相互作用

肠道内大量噬菌体是以温和性噬菌体状态存在的。实验室分离温和性噬菌体常用丝裂霉素 C

或紫外线激发宿主菌 SOS 保护机制释放温和性噬菌体^[79-80]。自然条件下,细菌可能受到人体摄入特殊食物成分等刺激,释放基因组内的温和性噬菌体。诱导出来的噬菌体可作为可移动遗传元件游离在肠道环境中,当其重新找到新的宿主回归溶源状态的时候,就可能带来不同菌株或不同菌种之间的基因重组,如抗性基因、毒力基因的扩散和交流^[81-82]。同时原宿主的裂解也可能带来肠道微生态的波动^[83]。此外,前噬菌体插入细菌基因组内特定部位可能会造成细菌整体的代谢改变,影响肠道环境组成。如在普通拟杆菌中,前噬菌体 BV01 插入位点位于胆汁酸盐水解酶相关基因 *TspO* 及其启动子之间,该噬菌体的存在抑制了普通拟杆菌对肠道环境中胆汁酸的降解^[84]。Oh 等通过给小鼠喂食短链脂肪酸,诱导了小鼠肠道内的罗伊氏乳杆菌前噬菌体的释放,促进了罗伊氏乳杆菌乙酸的分泌,并认为体内诱导该噬菌体可以提高该细菌在肠道内的竞争力^[79]。

4 肠道噬菌体在慢性疾病干预中的应用

4.1 肠道噬菌体在慢性疾病动物模型中的应用

随着肠道噬菌体的研究,研究者们开始探索应用肠道噬菌体来干预治疗肠道疾病或由肠道菌群引起的多样的慢性复杂疾病^[85]。有研究显示,在克罗恩病(Crohn's Disease, CD)患者的回肠黏膜上富集有一种粘附侵袭性大肠杆菌(Adherent Invasive *Escherichia coli*, AIEC), Galtier 等从污水中分离得到了 3 株可以靶向 AIEC 的噬菌体,并用该噬菌体鸡尾酒作用于表达 AIEC 受体的小鼠模型中,发现干预后的小鼠粪便和肠道部分的 AIEC 显著降低,并且单剂量的噬菌体鸡尾酒可以在 2 周内减少 AIEC 的定殖并预防肠炎症状的进展^[86-87]。

另外,利用噬菌体干预肠道外慢性疾病也获得一系列进展。Duan 等在酒精性肝炎患者的粪便样本分析结果中发现分泌细胞溶素的粪肠球菌与疾病严重度和死亡率高度相关,进而研究者从污水中

分离得到可靶向分泌细胞溶素粪肠球菌的噬菌体后,将其作用于定殖患者菌群的小鼠模型,小鼠肝脏细胞溶素水平降低,显著减轻了由酒精诱导的肝病表型^[88]。另外, Yuan 等在非酒精性脂肪性肝病 (Nonalcoholic Fatty Liver Disease, NAFLD) 患者的粪便样本中分离得到高酒精产力的肺炎克雷伯菌,并从环境中挖掘获得靶向该菌株的噬菌体,通过噬菌体靶向干预粪便样本后将其移植到小鼠体内,发现无明显 NAFLD 表征^[89]。与此同时,噬菌体也被尝试应用到干预治疗肿瘤疾病之中^[90-92]。Kabwe 等在对调控肿瘤微环境相关的微生物组的研究中,分离得到可靶向“促癌”性质的细菌的噬菌体,并将噬菌体调控作为靶向“促癌菌”的治疗方案进入到后续的研究当中^[91]。Zheng 等通过对公开数据库中结直肠癌和癌旁组织的样本信息分析,筛选出促结直肠癌的具核梭杆菌,并分离得到对具核梭杆菌有高特异靶向性噬菌体,并将该噬菌体 P2 与化疗药物连用,在仔猪和小鼠模型上均无明显不良反应,而且显示出一定效果的肿瘤消除作用^[90]。除此之外,针对由肠道菌群引起的复杂慢性疾病的致病目标菌种被越来越多地鉴定出来,这也为肠道噬菌体的应用提供更多的有效宿主菌资源,促进肠道噬菌体在慢性疾病中的应用^[4,93-94]。

4.2 肠道噬菌体在慢性疾病干预中的应用

目前国外已有公司针对多种慢性疾病干预开发的噬菌体产品。如成立于2015年的BiomX公司,该公司致力于使用天然和合成噬菌体定制噬菌体制剂,并将其应用到慢性疾病的治疗中。该公司在6年之内完成A轮和B轮的融资,共4亿人民币,并在纽约证券交易所和特拉维夫证券交易所上市。BiomX公司目前的研发管线针对干预炎症性肠病 (Inflammatory Bowel Disease, IBD)、原发性硬化性胆管炎 (Primary Sclerosing Cholangitis, PSC)、特异性皮炎、结直肠癌 (Colorectal cancer, CRC)、胃癌、痤疮、肝病等,并且多个项目已进入临床阶段 (<https://www.biomx.com/our-pipeline/>)。其中针对

痤疮的 BX001 是一款噬菌体鸡尾酒产品,是该公司的首条研发管线,通过靶向与痤疮相关的痤疮丙酸杆菌来调节皮肤菌群,改善皮肤痤疮,该管线目前已经进入临床 II 期试验;针对 IBD 的 BX002 也是一款噬菌体鸡尾酒产品,是根据 Atarashi 等的研究,锚定与 IBD 发作相关的细菌,其中一种是肺炎克雷伯菌,并针对这些细菌进行噬菌体药物的研发,该管线目前已经开展临床 I 期试验;而针对 PSC 和 CRC 的噬菌体产品也分别进入了临床 I 期和临床前期试验阶段;针对 IBD 和 PSC 的噬菌体产品在 I a 临床试验中取得阳性结果^[95]。另一家噬菌体公司 Locus Biosciences 针对炎症性肠病、免疫肿瘤药物不良反应、免疫检查点抑制剂相关感染以及结直肠癌等疾病展开相关噬菌体产品研发,并进入临床前期试验阶段 (<https://www.locus-bio.com/#pipeline>)。

5 总结与展望

随着宏基因组学的发展,越来越多的研究者投入到肠道菌群的研究中,肠道病毒组的研究也随之蓬勃发展。但病毒组学需基于已有的生物信息学技术进行分析,肠道中挑食性细菌菌株与肠道噬菌体颗粒的分离也进展缓慢,这些成为了肠道噬菌体研究的限速瓶颈。为了突破这些瓶颈,近年来研究者们不断探索,首先通过生物信息学的方法对样本中菌群的丰度进行分析,针对目标菌种丰度较高的样本进行目标菌种的筛选,积累肠道菌株资源^[18-19];针对肠道噬菌体的富集、核酸提取的方法进行优化^[96];使用针对性更强的独立数据库,不断积累和补充肠道噬菌体基因组信息,提升病毒序列预测软件等,即对不同样本中存在的不同的肠道噬菌体颗粒进行预测,并有针对性地靶向筛选肠道噬菌体颗粒,有效地提高获取肠道噬菌体颗粒的效率^[17,44];同时,尝试与宿主菌进行多轮共培养的方式对噬菌体颗粒进行富集,并在进行传统双琼脂平板滴定法之前,可以通过

qPCR等检测的方式对扩增后的噬菌体进行定量, 评估可获取噬菌体颗粒的几率^[23-24]。除此之外, 成立噬菌体研究联盟, 如目前深圳华大生命科学研究院等机构共建的全球噬菌体库(Global Phage Hub), 通过共享的形式搭建大规模、多元化的噬菌体实体和数据关联库, 有助于噬菌体领域越过技术和资源屏障得以迅猛发展。

目前肠道噬菌体相关的研究取得了一些积极的成果, 但未来仍然存在大量的机遇与挑战。

(1) 随着肠道噬菌体序列的不断积累, 关于肠道噬菌体的序列数据库与日俱增, 然而已公开的肠道噬菌体数据库规格不一、应用范围不同, 而且它们之间的相互交集未知, 导致在数据处理过程中易造成运算量冗余, 因此通过优化整合建立一个通用的肠道噬菌体的序列数据库势在必行。(2) 基于完整的肠道噬菌体序列数据库, 对肠道病毒组中的特异基因进行数据积累分析, 以用于定量分析肠道噬菌体组、识别噬菌体类型或判断肠道中噬菌体活性的工具开发。(3) 肠道菌群在不同疾病、不同人群中的表现具有差异性, 而病毒组与疾病相关的研究比较匮乏, 因此需要对病毒组在相应疾病中的表现进行研究, 并通过挖掘肠道病毒组中的标记物, 探索建立发病预警模型的可能性。(4) 不断增加的肠道噬菌体颗粒会提高肠道噬菌体基因组信息的多样性, 并可以赋予噬菌体基因组信息以外更多生物学特征信息。因此, 需要建立人体肠道噬菌体活体资源库, 并对基因组信息与噬菌体生物学特征以及其与宿主菌之间的互作关系进行深入研究, 进而扩充肠道噬菌体研究领域的基础。(5) 利用噬菌体干预复杂慢性疾病, 如代谢性疾病等, 通过分析患者的肠道菌群, 寻找慢性疾病的潜在致病菌株; 开展动物实验验证目标致病菌; 分析目标致病菌株的全基因组, 挖掘微生物致病因子; 进行目标致病菌噬菌体筛选, 开展动物模型噬菌体干预试验, 以期将肠道噬菌体作为新型干预慢性疾病的方法, 进而深入探索肠道噬菌体-宿主菌-人体免疫等的互作机制, 为

肠道噬菌体在复杂慢性疾病上的应用提供新的理论依据^[4,90]。

总之, 通过对肠道噬菌体序列数据的积累, 建立高效简洁的肠道噬菌体序列数据库; 开发具有定量、分类及活性预测功能的软件, 用于识别肠道噬菌体; 同时探索病毒组与疾病的相关性; 积累肠道噬菌体活体资源, 结合噬菌体基因组信息与生物学特征, 研究肠道噬菌体与宿主菌的相互作用; 并将肠道噬菌体活体颗粒应用到慢性疾病的干预治疗, 以完成肠道噬菌体由科研到应用的转化。

REFERENCES

- [1] Baquero F, Nombela C. The microbiome as a human organ[J]. *Clinical Microbiology and Infection*, 2012, 18: 2-4
- [2] Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body[J]. *PLoS Biology*, 2016, 14(8): e1002533
- [3] Li JH, Jia HJ, Cai XH, Zhong HZ, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, et al. An integrated catalog of reference genes in the human gut microbiome[J]. *Nature Biotechnology*, 2014, 32(8): 834-841
- [4] Liu RX, Hong J, Xu XQ, Feng Q, Zhang DY, Gu YY, Shi J, Zhao SQ, Liu W, Wang XK, et al. Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention[J]. *Nature Medicine*, 2017, 23(7): 859-868
- [5] Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota[J]. *Nature*, 2015, 528(7581): 262-266
- [6] Jie ZY, Xia HH, Zhong SL, Feng Q, Li SH, Liang SS, Zhong HZ, Liu ZP, Gao Y, Zhao H, et al. The gut microbiome in atherosclerotic cardiovascular disease[J]. *Nature Communications*, 2017, 8: 845
- [7] Yu J, Feng Q, Wong SH, Zhang DY, Liang QY, Qin YW, Tang LQ, Zhao H, Stenvang J, Li YL, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer[J]. *Gut*, 2017, 66(1): 70-78
- [8] Feng Q, Liang SS, Jia HJ, Stadlmayr A, Tang LQ, Lan Z, Zhang DY, Xia HH, Xu XY, Jie ZY, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence[J]. *Nature Communications*, 2015, 6: 6528
- [9] Zhang X, Zhang DY, Jia HJ, Feng Q, Wang DH, Liang D, Wu XN, Li JH, Tang LQ, Li Y, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly

- normalized after treatment[J]. *Nature Medicine*, 2015, 21(8): 895-905
- [10] Zhang YF, Gu YY, Ren HH, Wang SJ, Zhong HZ, Zhao XJ, Ma J, Gu XJ, Xue YM, Huang S, et al. Gut microbiome-related effects of berberine and probiotics on type 2 diabetes (the PREMOTe study)[J]. *Nature Communications*, 2020, 11: 5015
- [11] Shkoporov AN, Hill C. Bacteriophages of the human gut: the “known unknown” of the microbiome[J]. *Cell Host & Microbe*, 2019, 25(2): 195-209
- [12] Tao WY, Zhu S. Gut virome in health and disease[J]. *Journal of Biology*, 2019, 36(6): 1-5 (in Chinese)
陶万银, 朱书. 肠道病毒组与人体健康研究进展[J]. *生物学杂志*, 2019, 36(6): 1-5
- [13] Zhou B, Lin Y, Zhu WY. Advances in the interaction among gut phage, bacteria and host and its impacts on the animal host health[J]. *Chinese Journal of Animal and Veterinary Sciences*, 2019, 50(1): 14-20 (in Chinese)
周贝, 林焱, 朱伟云. 肠道噬菌体与细菌和宿主互动及其对动物机体健康影响的研究进展[J]. *畜牧兽医学报*, 2019, 50(1): 14-20
- [14] Wang Y, Chen Q. Intestinal phageome and human health[J]. *Journal of Microbes and Infections*, 2019, 14(5): 317-322 (in Chinese)
王悦, 陈倩. 肠道噬菌体组与人体健康[J]. *微生物与感染*, 2019, 14(5): 317-322
- [15] Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, Rohwer F. Metagenomic analyses of an uncultured viral community from human feces[J]. *Journal of Bacteriology*, 2003, 185(20): 6220-6223
- [16] Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data[J]. *PeerJ*, 2015, 3: e985
- [17] Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun FZ. VirFinder: a novel *k*-mer based tool for identifying viral sequences from assembled metagenomic data[J]. *Microbiome*, 2017, 5(1): 69
- [18] Zou YQ, Xue WB, Luo GW, Deng ZQ, Qin PP, Guo RJ, Sun HP, Xia Y, Liang SS, Dai Y, et al. 1, 520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses[J]. *Nature Biotechnology*, 2019, 37(2): 179-185
- [19] Forster SC, Kumar N, Anonye BO, Almeida A, Viciani E, Stares MD, Dunn M, Mkandawire TT, Zhu A, Shao Y, et al. A human gut bacterial genome and culture collection for improved metagenomic analyses[J]. *Nature Biotechnology*, 2019, 37(2): 186-192
- [20] Liu C, Du MX, Abuduaini R, Yu HY, Li DH, Wang YJ, Zhou N, Jiang MZ, Niu PX, Han SS, et al. Enlightening the taxonomy darkness of human gut microbiomes with a cultured biobank[J]. *Microbiome*, 2021, 9(1): 119
- [21] Hryckowian AJ, Merrill BD, Porter NT, Van Treuren W, Nelson EJ, Garlena RA, Russell DA, Martens EC, Sonnenburg JL. *Bacteroides thetaiotaomicron*-infecting bacteriophage isolates inform sequence-based host range predictions[J]. *Cell Host & Microbe*, 2020, 28(3): 371-379
- [22] Porter NT, Hryckowian AJ, Merrill BD, Fuentes JJ, Gardner JO, Glowacki RWP, Singh S, Crawford RD, Snitkin ES, Sonnenburg JL, et al. Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in *Bacteroides thetaiotaomicron*[J]. *Nature Microbiology*, 2020, 5(9): 1170-1181
- [23] Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, Hill C. Φ CrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*[J]. *Nature Communications*, 2018, 9: 4781
- [24] Guerin E, Shkoporov AN, Stockdale SR, Comas JC, Khokhlova EV, Clooney AG, Daly KM, Draper LA, Stephens N, Scholz D, et al. Isolation and characterisation of Φ crAss002, a crAss-like phage from the human gut that infects *Bacteroides xylanisolvens*[J]. *Microbiome*, 2021, 9(1): 89
- [25] Shkoporov AN, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, McDonnell SA, Khokhlova EV, Draper LA, Forde A, et al. The human gut virome is highly diverse, stable, and individual specific[J]. *Cell Host & Microbe*, 2019, 26(4): 527-541
- [26] Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB. The gut virome database reveals age-dependent patterns of virome diversity in the human gut[J]. *Cell Host & Microbe*, 2020, 28(5): 724-740
- [27] Ma YF, You XY, Mai GQ, Tokuyasu T, Liu CL. A human gut phage catalog correlates the gut phageome with type 2 diabetes[J]. *Microbiome*, 2018, 6(1): 24
- [28] Minot S, Sinha R, Chen J, Li HZ, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. The human gut virome: inter-individual variation and dynamic response to diet[J]. *Genome Research*, 2011, 21(10): 1616-1625
- [29] Zuo T, Sun Y, Wan YT, Yeoh YK, Zhang F, Cheung CP, Chen N, Luo J, Wang W, Sung JJY, et al. Human-gut-DNA virome variations across geography, ethnicity, and urbanization[J]. *Cell Host & Microbe*, 2020, 28(5): 741-751
- [30] Seo SU, Kweon MN. Virome-host interactions in intestinal health and disease[J]. *Current Opinion in Virology*, 2019, 37: 63-71
- [31] Liang GX, Zhao CY, Zhang HJ, Mattei L, Sherrill-Mix S, Bittinger K, Kessler LR, Wu GD, Baldassano RN, DeRusso P, et al. The stepwise assembly of the neonatal virome is modulated by breastfeeding[J]. *Nature*, 2020, 581(7809): 470-474
- [32] Bäckhed F, Roswall J, Peng YQ, Feng Q, Jia HJ, Kovatcheva-Datchary P, Li Y, Xia Y, Xie HL, Zhong HZ, et al. Dynamics and stabilization of the human gut microbiome during the first year of life[J]. *Cell Host & Microbe*, 2015, 17(5): 690-703

- [33] Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, Ryan FJ, Draper LA, Plevy SE, Ross RP, et al. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease[J]. *Cell Host & Microbe*, 2019, 26(6): 764-778
- [34] Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity[J]. *Cell*, 2021, 184(4): 1098-1109.e9
- [35] Benler S, Yutin N, Antipov D, Rayko M, Shmakov S, Gussow AB, Pevzner P, Koonin EV. Thousands of previously unknown phages discovered in whole-community human gut metagenomes[J]. *Microbiome*, 2021, 9(1): 78
- [36] Chen SF, Zhou YQ, Chen YR, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor[J]. *Bioinformatics*, 2018, 34(17): i884-i890
- [37] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30(15): 2114-2120
- [38] Chen YX, Chen YS, Shi CM, Huang ZB, Zhang Y, Li SK, Li Y, Ye J, Yu C, Li Z, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data[J]. *GigaScience*, 2018, 7(1): gix120
- [39] Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler[J]. *Genome Research*, 2017, 27(5): 824-834
- [40] Li DH, Luo RB, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices[J]. *Methods*, 2016, 102: 3-11
- [41] Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. CD-HIT: accelerated for clustering the next-generation sequencing data[J]. *Bioinformatics*, 2012, 28(23): 3150-3152
- [42] Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. MetaPhlan2 for enhanced metagenomic taxonomic profiling[J]. *Nature Methods*, 2015, 12(10): 902-903
- [43] Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2[J]. *Genome Biology*, 2019, 20(1): 257
- [44] Guo JR, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses[J]. *Microbiome*, 2021, 9(1): 37
- [45] Zheng TT, Li J, Ni YQ, Kang K, Misiakou MA, Imamovic L, Chow BKC, Rode AA, Bytzer P, Sommer M, et al. Mining, analyzing, and integrating viral signals from metagenomic data[J]. *Microbiome*, 2019, 7(1): 42
- [46] Zhao GY, Wu G, Lim ES, Droit L, Krishnamurthy S, Barouch DH, Virgin HW, Wang D. VirusSeeker, a computational pipeline for virus discovery and virome composition analysis[J]. *Virology*, 2017, 503: 21-30
- [47] Roux S, Tournayre J, Mahul A, Debroas D, Enault F. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis[J]. *BMC Bioinformatics*, 2014, 15: 76
- [48] Song WC, Sun HX, Zhang C, Cheng L, Peng Y, Deng ZQ, Wang D, Wang Y, Hu M, Liu W, et al. Prophage Hunter: An integrative hunting tool for active prophages[J]. *Nucleic Acids Research*, 2019, 47(W1): W74-W80
- [49] Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang YJ, Wishart DS. PHASTER: a better, faster version of the PHAST phage search tool[J]. *Nucleic Acids Research*, 2016, 44(W1): W16-W21
- [50] Nayfach S, Camargo AP, Schulz F, Eloie-Fadrosch E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes[J]. *Nature Biotechnology*, 2021, 39(5): 578-585
- [51] Bolduc B, Jang HB, Doucier G, You ZQ, Roux S, Sullivan MB. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect *Archaea* and *Bacteria*[J]. *PeerJ*, 2017, 5: e3243
- [52] Jang HB, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks[J]. *Nature Biotechnology*, 2019, 37(6): 632-639
- [53] Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. CRISPRCasFinder, an update of CRISPRfinder, includes a portable version, enhanced performance and integrates search for Cas proteins[J]. *Nucleic Acids Research*, 2018, 46(W1): W246-W251
- [54] Zhang F, Zhao SJ, Ren CY, Zhu YW, Zhou HB, Lai YK, Zhou FX, Jia YQ, Zheng KJ, Huang ZW. CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions[J]. *Communications Biology*, 2018, 1: 180
- [55] Li ML, Wang YN, Li FY, Zhao Y, Liu MY, Zhang SJ, Bin YN, Smith AI, Webb G, Li J, et al. A deep learning-based method for identification of bacteriophage-host interaction[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020(99): 1
- [56] Lu CY, Zhang Z, Cai ZN, Zhu ZZ, Qiu Y, Wu AP, Jiang TJ, Zheng HP, Peng YS. Prokaryotic virus host predictor: A Gaussian model for host prediction of prokaryotic viruses in metagenomics[J]. *BMC Biology*, 2021, 19(1): 5
- [57] Nooij S, Schmitz D, Vennema H, Kroneman A, Koopmans MPG. Overview of virus metagenomic classification methods and their biological applications[J]. *Frontiers in Microbiology*, 2018, 9: 749
- [58] Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li JH, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes[J]. *Nature Biotechnology*,

- 2014, 32(8): 822-828
- [59] Sutton TDS, Clooney AG, Ryan FJ, Paul Ross R, Hill C. Choice of assembly software has a critical impact on virome characterisation[J]. *bioRxiv*, 2018. DOI: 10.1101/479105
- [60] Roux S, Páez-Espino D, Chen IMA, Palaniappan K, Ratner A, Chu K, Reddy TBK, Nayfach S, Schulz F, Call L, et al. IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses[J]. *Nucleic Acids Research*, 2021, 49(D1): D764-D775
- [61] O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation[J]. *Nucleic Acids Research*, 2016, 44(D1): D733-D745
- [62] Marbouty M, Thierry A, Koszul R. Phages-bacteria interactions network of the healthy human gut[J]. *bioRxiv*, 2020. DOI: 10.1101/2020.05.13.093716
- [63] Alawi M, Burkhardt L, Indenbirken D, Reumann K, Christopheit M, Kröger N, Lütgehetmann M, Aepfelbacher M, Fischer N, Grundhoff A. DAMIAN: an open source bioinformatics tool for fast, systematic and cohort based analysis of microorganisms in diagnostic samples[J]. *Scientific Reports*, 2019, 9: 16841
- [64] Guerin E, Hill C. Shining light on human gut bacteriophages[J]. *Frontiers in Cellular and Infection Microbiology*, 2020, 10: 481
- [65] Qin XY, Yang HJ. Research progress on techniques for separation, purification of bacteriophages[J]. *China Biotechnology*, 2020, 40(5): 78-83 (in Chinese)
秦旭颖, 杨洪江. 噬菌体分离纯化技术研究进展[J]. *中国生物工程杂志*, 2020, 40(5): 78-83
- [66] Soleimani-Delfan A, Bouzari M, Wang R. A rapid competitive method for bacteriophage genomic DNA extraction[J]. *Journal of Virological Methods*, 2021, 293: 114148
- [67] Hoyles L, McCartney AL, Neve H, Gibson GR, Sanderson JD, Heller KJ, van Sinderen D. Characterization of virus-like particles associated with the human faecal and caecal microbiota[J]. *Research in Microbiology*, 2014, 165(10): 803-812
- [68] Sutton TDS, Hill C. Gut bacteriophage: current understanding and challenges[J]. *Frontiers in Endocrinology*, 2019, 10: 784
- [69] Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, et al. Minimum information about an uncultivated virus genome (MIUViG)[J]. *Nature Biotechnology*, 2019, 37(1): 29-37
- [70] Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling LC, Barr JJ, Speth DR, Seguritan V, Aziz RK, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes[J]. *Nature Communications*, 2014, 5: 4498
- [71] Devoto AE, Santini JM, Olm MR, Anantharaman K, Munk P, Tung J, Archie EA, Turnbaugh PJ, Seed KD, Blekhan R, et al. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes[J]. *Nature Microbiology*, 2019, 4(4): 693-700
- [72] Hobbs Z, Abedon ST. Diversity of phage infection types and associated terminology: the problem with 'Lytic or lysogenic'[J]. *FEMS Microbiology Letters*, 2016, 363(7): fnw047
- [73] Mirzaei MK, Maurice CF. Ménage à trois in the human gut: interactions between host, bacteria and phages[J]. *Nature Reviews Microbiology*, 2017, 15(7): 397-408
- [74] Silveira CB, Rohwer FL. Piggyback-the-Winner in host-associated microbial communities[J]. *NPJ Biofilms and Microbiomes*, 2016, 2: 16010
- [75] Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, Coutinho FH, Dinsdale EA, Felts B, Furby KA, et al. Lytic to temperate switching of viral communities[J]. *Nature*, 2016, 531(7595): 466-470
- [76] Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet Masters of the marine microbial realm[J]. *Nature Microbiology*, 2018, 3(7): 754-766
- [77] Lourenço M, Chaffringeon L, Lamy-Besnier Q, Pédrón T, Campagne P, Eberl C, Bérard M, Stecher B, Debarbieux L, De Sordi L. The spatial heterogeneity of the gut limits predation and fosters coexistence of bacteria and bacteriophages[J]. *Cell Host & Microbe*, 2020, 28(3): 390-401.e5
- [78] Jiang XF, Hall AB, Arthur TD, Plichta DR, Covington CT, Poyet M, Crothers J, Moses PL, Tolonen AC, Vlamakis H, et al. Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut[J]. *Science*, 2019, 363(6423): 181-187
- [79] Oh JH, Lin XB, Zhang SW, Tollenaar SL, Özçam M, Dunphy C, Walter J, Van Pijkeren JP. Prophages in *Lactobacillus reuteri* are associated with fitness trade-offs but can increase competitiveness in the gut ecosystem[J]. *Applied and Environmental Microbiology*, 2019. DOI: 10.1128/aem.01922-19
- [80] Kreuzer KN. DNA damage responses in prokaryotes: Regulating gene expression, modulating growth patterns, and manipulating replication forks[J]. *Cold Spring Harbor Perspectives in Biology*, 2013, 5(11): a012674
- [81] Manrique P, Dills M, Young M. The human gut phage community and its implications for health and disease[J]. *Viruses*, 2017, 9(6): 141
- [82] Hu Q, Liu C, Zhang D, Wang R, Qin LL, Xu Q, Che LQ, Gao F. Effects of low-dose antibiotics on gut immunity and antibiotic resistomes in weaned piglets[J]. *Frontiers in Immunology*, 2020, 11: 903
- [83] Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L,

- Silver PA, Gerber GK. Dynamic modulation of the gut microbiota and metabolome by bacteriophages in a mouse model[J]. *Cell Host & Microbe*, 2019, 25(6): 803-814
- [84] Campbell DE, Ly LK, Ridlon JM, Hsiao A, Whitaker RJ, Degnan PH. Infection with *Bacteroides* phage BV01 alters the host transcriptome and bile acid metabolism in a common human gut microbe[J]. *Cell Reports*, 2020, 32(11): 108142
- [85] Wahida A, Tang F, Barr JJ. Rethinking phage-bacteria-eukaryotic relationships and their influence on human health[J]. *Cell Host & Microbe*, 2021, 29(5): 681-688
- [86] Galtier M, Sordi LD, Sivignon A, De Vallée A, Maura D, Neut C, Rahmouni O, Wannerberger K, Darfeuille-Michaud A, Desreumaux P, et al. Bacteriophages targeting adherent invasive *Escherichia coli* strains as a promising new treatment for Crohn's disease[J]. *Journal of Crohn's and Colitis*, 2017, 11(7): 840-847
- [87] Small CLN, Reid-Yu SA, McPhee JB, Coombes BK. Persistent infection with Crohn's disease-associated adherent-invasive *Escherichia coli* leads to chronic inflammation and intestinal fibrosis[J]. *Nature Communications*, 2013, 4: 1957
- [88] Duan Y, Llorente C, Lang S, Brandl K, Chu HK, Jiang L, White RC, Clarke TH, Nguyen K, Torralba M, et al. Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease[J]. *Nature*, 2019, 575(7783): 505-511
- [89] Yuan J, Chen C, Cui JH, Lu J, Yan C, Wei X, Zhao XN, Li NN, Li SL, Xue GH, et al. Fatty liver disease caused by high-alcohol-producing *Klebsiella pneumoniae*[J]. *Cell Metabolism*, 2019, 30(4): 675-688.e7
- [90] Zheng DW, Dong X, Pan P, Chen KW, Fan JX, Cheng SX, Zhang XZ. Phage-guided modulation of the gut microbiota of mouse models of colorectal cancer augments their responses to chemotherapy[J]. *Nature Biomedical Engineering*, 2019, 3(9): 717-728
- [91] Kabwe M, Dashper S, Bachrach G, Tucci J. Bacteriophage manipulation of the microbiome associated with tumour microenvironments-can this improve cancer therapeutic response?[J]. *FEMS Microbiology Reviews*, 2021. DOI: doi.org/10.1093/femsre/fuab017
- [92] Dong X, Pan P, Zheng DW, Bao P, Zeng X, Zhang XZ. Bioinorganic hybrid bacteriophage for modulation of intestinal microbiota to remodel tumor-immune microenvironment against colorectal cancer[J]. *Science Advances*, 2020, 6(20): eaba1590
- [93] Wong SH, Yu J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications[J]. *Nature Reviews Gastroenterology & Hepatology*, 2019, 16(11): 690-704
- [94] Zhu F, Ju YM, Wang W, Wang Q, Guo RJ, Ma QY, Sun Q, Fan YJ, Xie YY, Yang Z, et al. Metagenome-wide association of gut microbiome features for schizophrenia[J]. *Nature Communications*, 2020, 11(1): 1612
- [95] Atarashi K, Suda W, Luo CW, Kawaguchi T, Motoo I, Narushima S, Kiguchi Y, Yasuma K, Watanabe E, Tanoue T, et al. Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation[J]. *Science*, 2017, 358(6361): 359-365
- [96] Yang FM, Sun JH, Luo HN, Ren HH, Zhou HC, Lin YX, Han M, Chen B, Liao HL, Brix S, et al. Assessment of fecal DNA extraction protocols for metagenomic studies[J]. *GigaScience*, 2020, 9(7): giaa071