参赛队员姓名：王兆琪

中学：北京师范大学附属实验中学

省份：北京市

国家/地区：中国

指导教师姓名：向烨，刘哲

指导教师单位：清华大学

论文题目：**一株肠道内丰富的溶木聚糖拟杆菌温和型噬菌体ΦHM7的分离与鉴定**

一株肠道内丰富的溶木聚糖拟杆菌温和型噬菌体ΦHM7 的分离与鉴定

作　者：王兆琪

摘要

肠道噬菌体作为肠道内最丰富的微生物在人类健康与疾病中发挥着至关重要的作用。由于分离肠道噬菌体相对困难，其实体资源库非常匮乏，严重限制了肠道噬菌体的分类及机制的研究，近年来肠道噬菌体的研究主要依赖于宏基因组分析，因此本研究通过分离鉴定并表征肠道温和型噬菌体，将是推动肠道菌群稳态研究的关键因素。

最近研究发现肠道拟杆菌*Bacteroides Xylanisolvens HM7*能够帮助降解尼古丁，从而缓解因尼古丁引起的脂肪肝。本研究通过系统的筛选，从粪便与污水样本中分离获得了*Bacteroides Xylanisolvens HM7*的噬菌体ΦHM7。通过全基因组分析发现ΦHM7为温和型噬菌体。电镜观察发现其大小约为60 nm的头部衣壳，以及长度约为200 nm的尾管。从形态分析其应属于长尾噬菌体科。通过计划关系发现ΦHM7与温和型噬菌体p00相似，且宏基因组丰度分析显示ΦHM7是人类粪便中丰度较高的噬菌体，且其在健康人群与患病人群中的丰度存在一定差异，因此ΦHM7可以作为影响健康的驱动因子存在。同时我们发现ΦHM7的基因组含有独特的多样性生成逆转录因子 (DGR)，其靶向具有独特的序列信息及结构特点的可变蛋白gp40。gp40可能是噬菌体潜在的尾丝蛋白出现。同时我们重组表达了可变蛋白gp40，为进一步的结构与功能研究奠定了新的基础。总结来说，我们分离鉴定并表征了一株广泛存在于人体肠道内的温和型噬菌体ΦHM7，且其独特的DGR系统可能指导了它广泛的侵袭多种宿主，这与我们所熟知的专一性噬菌体不同，其多宿主侵袭能力可能具有潜在的肠道菌群调节能力，未来可为进一步探究噬菌体调节肠道菌群及代谢物，以及肠道微生物的定向改造提供有力的工具。

关键词：肠道微生物 溶木聚糖拟杆菌噬菌体 温和型噬菌体 多样性生成逆转录因子 (DGR)

目录

第 1 章 研究背景

1.1 肠道微生物概述

肠道内具有复杂的微生物环境，他们在组成上存在一定的个体差异性。肠道微生物包括我们熟知的肠道细菌，此外还有真菌、病毒等微生物。近些年的研究表明肠道内的微生物影响着人们的健康。通过基因测序等宏基因组学等技术分析健康人群与患有一系列疾病的人群的肠道微生物的差异性加速了对肠道微生物的研究。目前针对肠道微生物的研究主要以细菌为主，研究表明肠道微生物有助于宿主的代谢途径 (Ghosh et al., 2022)，还具有免疫调节的功能 (Schiavi et al., 2016; Lebeer et al., 2018; Couto et al., 2020)。此外，一些研究表明肠道微生物与神经系统形成脑肠轴，通过肠道神经系统、微生物代谢物的循环作用等影响神经系统 (Burokas et al., 2015)。

1.2 肠道噬菌体

肠道微生物在胃肠道、神经系统等一系列疾病中起着至关重要的作用，然而这些微生物组的研究主要关注到细菌，而真菌、病毒等研究较少。然而肠道微生物中噬菌体作为含量最多的病毒(Reyes et al., 2010; Minot et al., 2011)，可能通过影响其宿主菌来调节人类健康(Camarillo-Guerrero et al., 2021)。

1.2.1 肠道噬菌体概述

噬菌体作为一类以细菌作为宿主的病毒于 20 世纪初首次由 d’Hérelle 和 Twort 发现，目前对于噬菌体的研究主要集中在替代传统抗生素治疗来面对耐药菌的感染，而噬菌体的鉴定方法主要依赖于分离培养与宏基因组学等，通过一定的培养方法将复杂环境中的病毒颗粒分离出来或通过将环境、粪便样品进行基因组分析来组建基因文库进行研究。但是噬菌体在肠道内作用的研究相对较少，由于多数肠道微生物在体外难以培养，目前分离到的噬菌体多数来源于陆地与湖泊，只有少数的噬菌体从人类粪便分离得到。Stephan J. Ott 等人利用无菌的 FFT 移植，成功的恢复了由抗生素破坏的肠道微生物群，首次证明肠道微生物中除了细菌以外，噬菌体可以作为潜在的肠道内环境稳态调节剂 (Ott et al., 2017)，由此可知肠道噬菌体作为研究对象具有巨大的潜力。

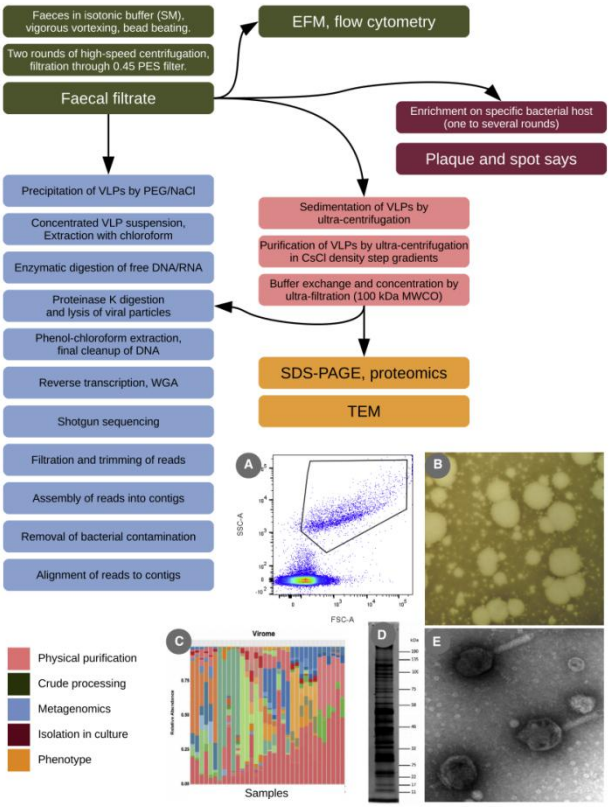

图 1.1 研究肠道噬菌体的工作流程图(Shkoporov and Hill, 2019)

图中显示了宏基因组分析、体外培养、电子显微镜观察、蛋白质组学等方法的流程。

目前对于肠道噬菌体的研究主要依赖于基因测序、宏基因组学、体外病毒颗粒富集培养等方法，如图 1.1 所示。通过电镜观察发现，多数已观察到的粪便样品噬菌体属于长尾噬菌体科（*Siphoviridae*），肌尾噬菌体科（*Myoviridae*）和短尾噬菌体科（*Podoviridae*），如图 1.2 所示，通过分析粪便中病毒的宏基因组测序结果表明其中 99%的序列与数据库中已知噬菌体序列无同源性，且多数的序列是不完整的 (Aggarwala et al., 2017)。

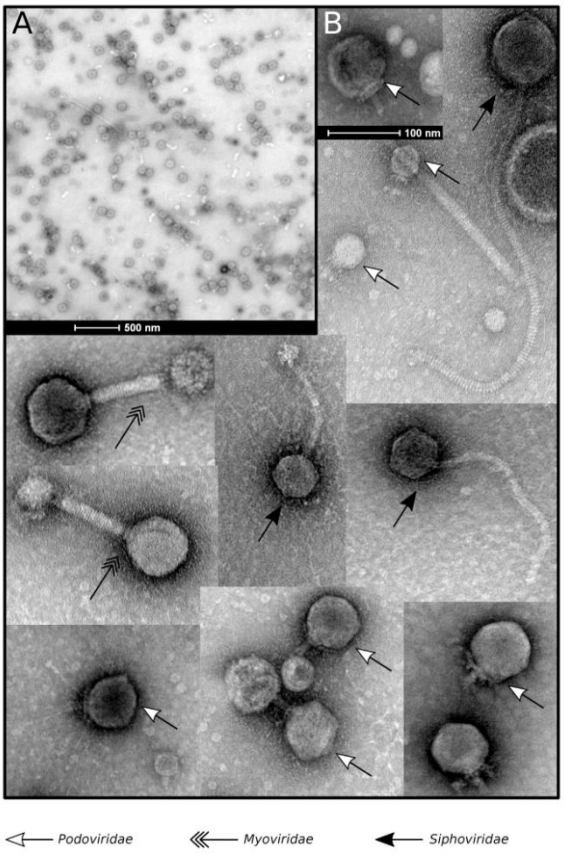

图 1.2 透射电子显微镜下观察人类粪便样品中噬菌体的主要形态(Shkoporov and Hill, 2019)

1.2.2 噬菌体分类及宿主特异性

根据国际病毒委员会(International Committee on Taxonomy of Viruses，ICTV) 的分类标准，可以通过噬菌体的遗传物质进行分类，包括 ssDNA、dsDNA、ssRNA 和 dsRNA 噬菌体(Adams et al., 2017; Dion et al., 2020)。这些噬菌体具有不同的遗传物质组成且在形态结构上存在明显的差异性，如图 1.3 所示。其中 ssDNA 噬菌体主要包括微小噬菌体科（*Microviridae*）与丝状噬菌体科（*Inoviridae*），微小噬菌体科的代表噬菌体为 phiX174(Sun et al., 2014)，丝状噬菌体科的代表噬菌体为 M13(Marvin et al., 2014)。对于 dsDNA 噬菌体主要包括有尾与无尾两类，其中绝大多数的 dsDNA 噬菌体为有尾噬菌体，包括肌尾噬菌体科（*Myoviridae*）、短尾噬菌体科（*Podoviridae*）、埃克曼噬菌体科（*Ackermannviridae*）和长尾噬菌体科（*Siphoviridae*）。这几类代表噬菌体分别为 T4 噬菌体、T7 噬菌体、AG3 噬菌体以及λ噬菌体 (Hohn and Katsura, 1977)。

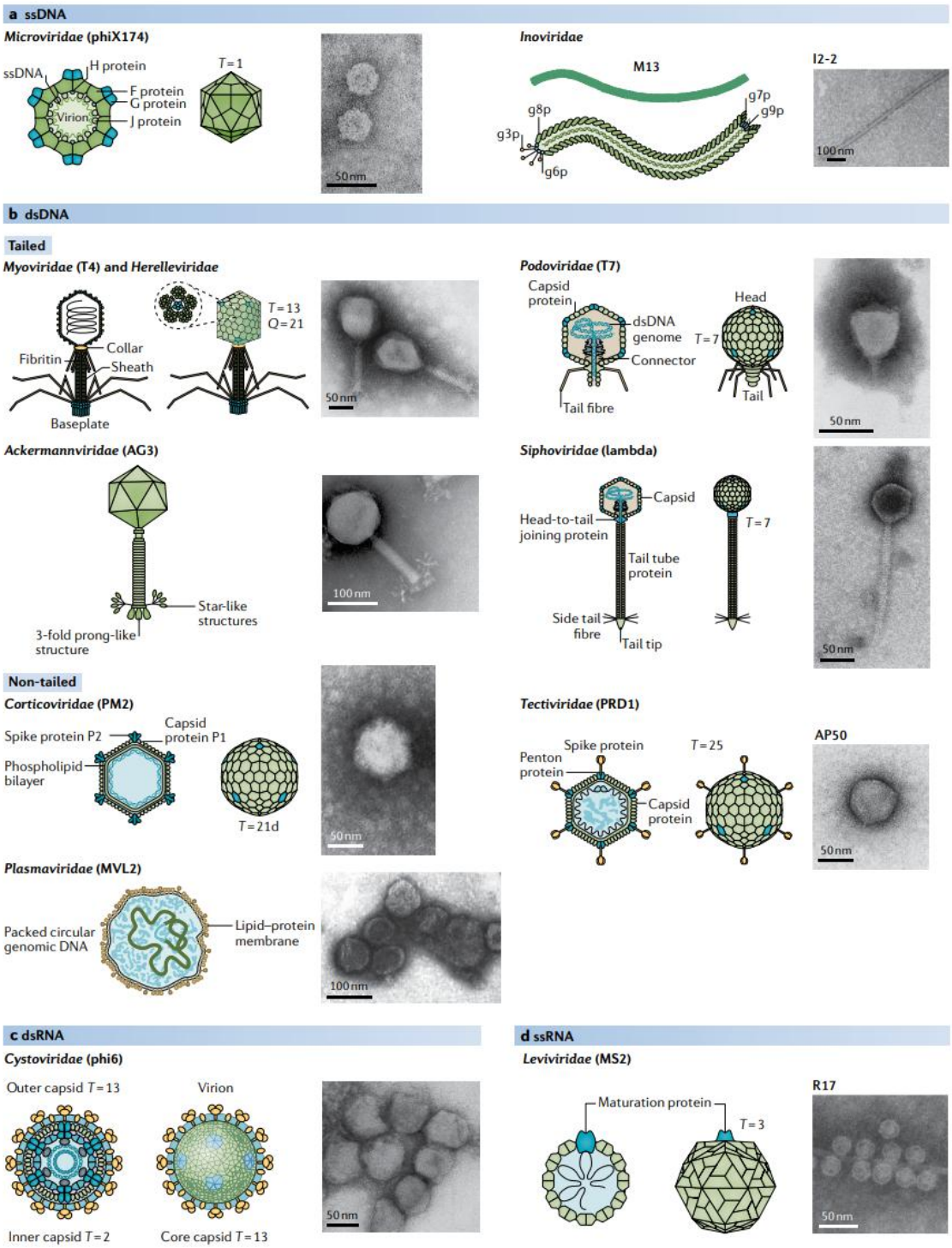

图 1.3 噬菌体的形态学和基因组类型分类(Dion et al., 2020)

图中展示了噬菌体在遗传物质，尾部形态，蛋白质衣壳的多样性。

1.2.3 温和型噬菌体Φp00–Hankyphage

2018年，Forest Rohwer团队在挖掘DGR系统的过程中发现了Φp00–Hankyphage，Hankyphage存在于*Bacteroides dorei* 拟杆菌中(Benler et

al., 2018)，如图1.4所示。Hankyphage的基因组长度为43kbp，如图1.5所示。在13个拟杆菌种属中均出现相似基因，相似度达95%以上，如图1.6。对其进行噬菌斑实验发现Hankyphage在这些拟杆菌里不产生噬菌斑，且含有独特的Mu-like转座机制，在宿主基因组上的整合位点是随机的。而在Hankyphage上的DGR系统的VR区所在蛋白功能是未知的 (Benler et al., 2018)。

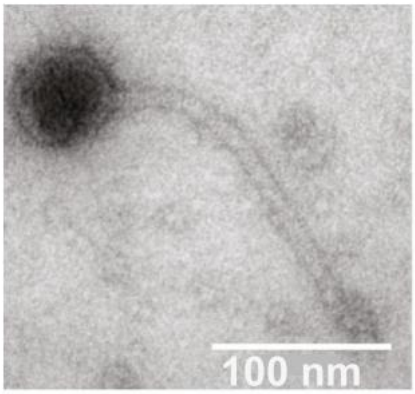

图 1.4 噬菌体Φp00–Hankyphage 的形态 (Benler et al., 2018; Dion et al., 2020)

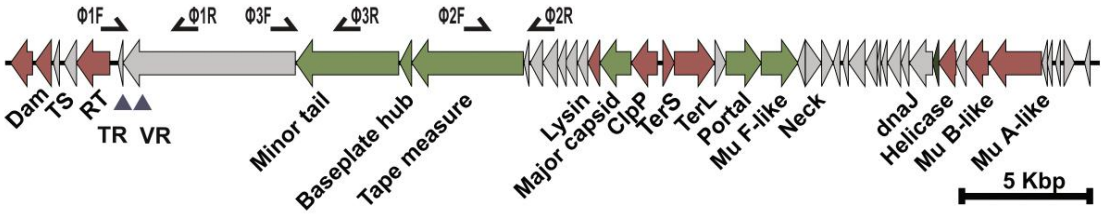

图 1.5 噬菌体Φp00–Hankyphage 的基因组图谱示意图 (Benler et al., 2018; Dion et al., 2020)

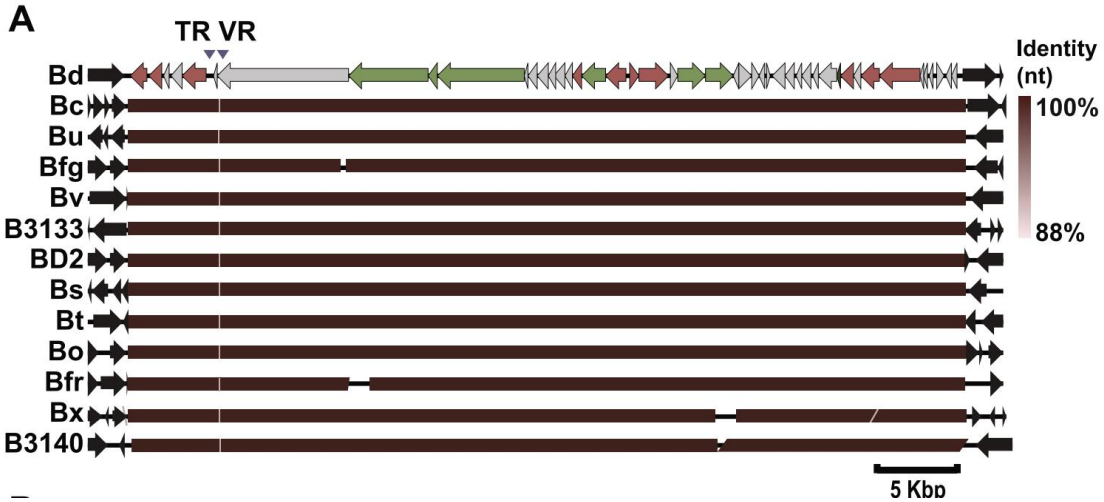

图 1.6 噬菌体Φp00–Hankyphage 存在于多种拟杆菌基因组中 (Benler et al., 2018)

1.3 多样性生成逆转录因子 (DGR)

1.3.1 DGR 系统概述

多样性生成逆转录元件（DGR）属于逆转录因子家族，其能够通过快速诱变编码蛋白质的靶基因来加速噬菌体的进化(Liu et al., 2002; Doulatov et al., 2004)，从而使噬菌体获得广泛的宿主侵染能力。DGR 系统最初是在博德特氏菌属噬菌体中被发现的 (Liu et al., 2002)，目前在细菌、古细菌及其病毒中发现了近 60,000 个 DGR 系统 (Paul et al., 2017; Roux et al., 2021)。DGR 系统通过诱变性回巢的、独特的、单向的、复制和替换机制来发挥功能，其需要至少三个组件：包括含有作为诱变受体的可变重复序列（VR）的靶基因、与 VR 几乎相同的不变区模板重复序列（TR）和 DGR 编码的逆转录酶（RT），如图 1.7 所示。DGR 系统靶标蛋白的多样性通过 TR、VR 之间的位点特异性机制产生，其中对应于 TR 中含腺嘌呤密码子的 VR 残基被选择性诱变 (Liu et al., 2002) 如图 1.7。在靶标蛋白多样化过程中，TR 衍生的 RNA 中间体作为通过易错的 DGR 编码 RT 合成 cDNA 的模板，在逆转录过程中腺嘌呤残基不成比例地错配，导致四种 dNTP 中的任何一种在特定位点随机掺入 cDNA 中 (Guo et al., 2008)。随后，所得的腺嘌呤诱变的 cDNA 替换靶基因内的 VR 序列，产生在特定结构域中的指定位点多样化的可变蛋白。因为 TR 在这个过程中没有改变，VR 可以经历重复的回巢和自然选择以进化蛋白质功能。

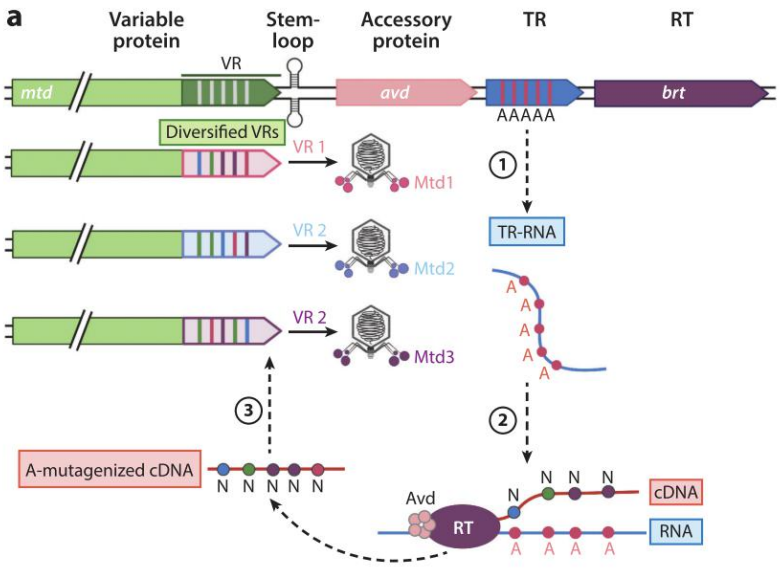

图 1.7 噬菌体 DGR 系统工作示意图 (Macadangdang et al., 2022)

1.3.2 DGR 系统可变区蛋白的结构与功能

DGR 系统可变区所在蛋白具有广泛的多样性，且目前近 70%的可变区蛋白缺乏功能信息的注释(Lobb et al., 2020; Roux et al., 2021)，绝大多数的生理作用仍不清楚。尽管如此，已经在可变蛋白质中鉴定的具有可预测功能的基序的分类表明 DGR 定向诱变在微生物世界中具有广泛的实用性。最广泛表征的 DGR 可变蛋白是 BPP-1 尾部纤维受体结合蛋白 Mtd (Liu et al., 2002; McMahon et al., 2005; Dai et al., 2010)。Mtd 作为 DGR 系统靶标的蛋白，其结构和功能研究揭示了通常在可变蛋白质中可能保守的结构构象，如图 1.8 所示。Mtd 指导决定噬菌体的结合特异性，其多样化由 VR 内的多个非同义氨基酸取代产生，且都特异性地发生在其 C 型凝集素（C-Lec）折叠的结合口袋内(McMahon et al., 2005)。构成配体结合口袋残基的选择性多样化驱动 Mtd 特异性的进化，使得噬菌体可以携带不同配体结合能力的尾纤维蛋白 (Miller et al., 2008)。

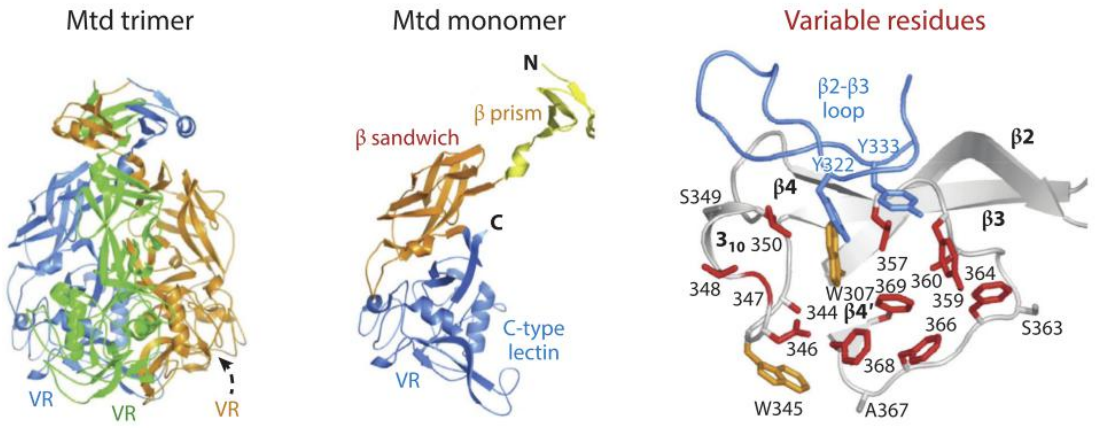

图 1.8 噬菌体 DGR 系统可变蛋白 Mtd 的结构 (Macadangdang et al., 2022)

1.4 噬菌体尾部结构概述

长尾噬菌体的尾部可以分为非收缩和收缩两种类型，尾部上端是尾部接头蛋白，尾部主体由尾管蛋白 TTP 组成，末端是尾尖复合体 TTC，与宿主细胞直接相互作用完成侵染，由于噬菌体组装过程是从 TTC 开始组装，TTP 以 TTC 作为基板向上组装并连接至尾部接头蛋白，因此尾部接头蛋白也称做尾部终止蛋白 TrP。不同噬菌体的 TTC 呈现出不同的形状，TTC 表面存在两种尾部纤维蛋白，中央纤维蛋白 CFP 和侧纤维蛋白 SFP，如图 1.9 所示。

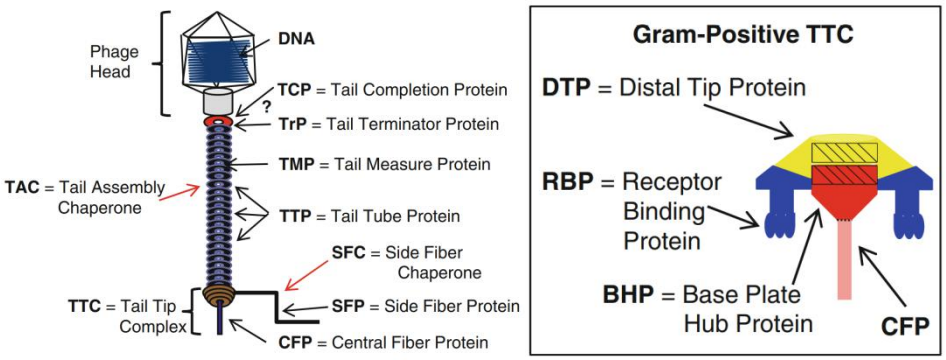

图 1.9 长尾噬菌体的尾部保守特征及相关尾部蛋白(Davidson et al., 2012)

第 2 章 材料与方法

2.1 *Bacteroides Xylanisolvens HM7* 的分离与培养

（1）粪便样品收集及保存：收集新鲜粪便样品存于管内放置于厌氧操作台内。在厌氧操作台中对粪便进行培养，剩余样品在厌氧操作台内分装并存于-80℃冰箱。

（2）肠道细菌的分离培养：采用直接涂板的方式，取约 1 g 粪便用 1 ml 生理盐水进行稀释，并于平板划线。培养方式选用 37 ℃厌氧培养 24 h，挑取平板上的菌落在 BHI 液体培养基（5 ml）中分别于有氧和厌氧环境中 37 ℃富集培养。

（3）菌株鉴定：取平板上的单克隆接种与 BHI 培养基中扩增，将扩增后的菌液委托睿博测序公司进行序列纯化和双向测序，其中 PCR 采用通用引物 27F:5’-AGTCTCTGATCATGCCTCAG-3’ 和 1492R:5’-AAGGAGGTGCTCCAGCC-3’对菌株的基因组 DNA 进行扩增。将测序结果提交至 NCBI 利用 Blastn 进行序列比对，鉴定细菌的种水平。选取其中的 *Bacteroides Xylanisolvens HM7* 进行进一步的实验。

2.2 ΦHM7 噬菌体的分离与培养

（1）噬菌体的分离和筛选：将污水样品 4000 × g，4 ℃离心 10 min，取 49 mL 上清和 1 mL 宿主菌加至 50 mL 2 × LB 液体培养基中，置于 37 ℃培养 48 h 左右，将混合培养液 8000 × g，4 ℃离心 10 min，去除菌体和杂质碎片。最后，上清依次经过 0.45 μm 滤膜过滤得到噬菌体原液。

（2）点滴法筛选裂解性噬菌体：将宿主菌新鲜菌液均匀涂布于 LB 固体平板上；待平板晾干后，吸取 5 μL-7 μL 噬菌体原液滴至平板上；待液滴晾干后，倒置于 37 ℃恒温培养箱内；过夜培养，观察平板上是否出现透明的噬菌斑。

（3）噬菌体富集：

噬菌体的富集过程如下：首先，挑取纯化的噬菌斑，吹入 200 μL TMS 缓冲液中，重复吹打均匀后，取 100 μL 混入 1 mL 处于对数期的菌液内，室温静置 10 min 后，37 ℃振荡培养后得到清亮透光的噬菌体裂解液；8000 × g 离心，使用梯度扩增依次加入 5 ml、20 ml、50 ml 菌液内。将梯度扩增液 8000 × g，4 ℃离心 10 min，再将上清液 0.45 μm 滤膜过滤，得到噬菌体富集液。该富集液可在 4 ℃长期保藏。将富集液 7,8000 × g 离心 2 h，去掉上清液，使用 TMS 缓冲液清洗 2-3 次，加入 1-2 ml TMS 缓冲液 4 ℃静置 12 h 后重悬并保存，该浓缩液即噬菌体原液可在 4 ℃长期保藏。

2.3 噬菌体基因组的提取

从噬菌体浓缩液中提取并纯化噬菌体基因组。将 20 μL 浓缩的噬菌体与 500 μL TE 缓冲液（50 mM Tris pH 7.8，10 mM EDTA）和蛋白酶 K（6 μL/500 μL）混合，在 70℃下孵育 1 h。然后加入 500 μL 25:24:1 的苯酚/氯仿/异丙醇，轻轻倒置，充分混合。将溶液在 12,000 xg 离心 5 min，将约 250 μL 水相上清液小心地转移到一个新的管中。加入 750 μL 预冷的无水乙醇后-80℃放置 12 h，然后 12,000 xg 10 min，并用 75%乙醇冲洗两次。最后用 50 μL ddH2O 重新悬浮基因组 DNA，将得到的噬菌体基因组委托赛默百合公司进行全基因组序列测序。

2.4 负染样品制备

（1）染色液配制：使用双蒸馏水（ddH_2O）配制 2%醋酸铀溶液（PH：4.5）, 4 ℃避光保存。

（2）铜网亲水化处理：将铜网置于等离子清洗仪中，真空 2 min，中等离子强度轰击 40 s，使其变得亲水，能够吸附蛋白质与染色剂。

（3）蛋白质负染色：首先将染料 12000 rpm 离心 10 min 去掉杂质，用镊子夹取亲水化处理的铜网载体，使用移液器吸取 7 μL 样品加置铺有碳膜一面，静置 1 min 后用滤纸吸取。然后吸取 5 μL 染色液置于碳膜一面并用滤纸快速吸取，重复 2 次。最后吸取 5 μL 染色液置于碳膜一面静置 1 min 吸干、风干即可。

2.5 基因组分析

（1）blast 分析：将ΦHM7 基因组在 NCBI 数据库中进行 blastn 查询，将比对结果中所有注释完整的细菌基因组下载下来，提取其中的温和噬菌体序列。再利用 Blastp 比对噬菌体末端酶大亚基序列。

（2）进化树构建：利用 MEGA-X 软件对 39 条ΦHM7 同源末端酶大亚基进行多序列比对，多序列比对选择 ClustalW 模式，构建进化树选择 NJ 树，最后提交到 iTOL 网站 (https://itol.embl.de/) 进行数据可视化。

（3）宏基因组分析：在人类统一肠道病毒组目录 UHGV 数据库 (https://portal.nersc.gov/UHGV/) 中查询ΦHM7 基因组序列，利用数据库已有信息查看其在宏基因组、宏病毒组中的分布情况，下载数据用 EXCEL 归纳整理。数据提交到 Hiplot 网站(https://hiplot.com.cn/home/index.html)选取柱状图、热图进行相应图片绘制。

2.6 分子克隆

本研究中潜在尾纤维蛋白 gp40 的表达选用原核表达体系，所用克隆载体为 pET28a，所带标签为 N 端 6xHis，分子克隆构建的基本流程如下：

（1）引物设计：利用目的片段序列信息设计片段线性化 3‘端及 5‘端片段引物，并在其两端加上与 pET28a 酶切位点序列。

（2）载体线性化：利用 PCR 将 pET28a 进行线性化，并用 DpnI将模板消化。

（3）目的片段扩增：利用ΦHM7 全基因组序列，将片段引物通过 PCR 将目的片段线性化并用 DpnI将模板消化。

（4）胶回收：对以上 PCR 产物进行琼脂糖凝胶电泳，根据 DNA Maker 确定目的条带的大小，将目的条带切下并使用 Axygen 试剂盒回收产物。

（5）无缝克隆：使用 Infusion 试剂盒进行无缝连接，恒温 37℃孵育 30min

（6）转化：将克隆产物转化扩增质粒 DH5α中，冰上静置 30min，42℃热激 90 s，冰上静置 5 min，37℃活化并涂布于平板。

（7）测序：挑取数个平板上单克隆，将其加入 2 ml LB 培养基中 37℃培养 3-4 h，利用 Axygen 试剂盒提取质粒并送公司测序。

2.7 蛋白的纯化与表达

1.表达质粒的转化

将测序正确的克隆质粒转化表达质粒 BL21 中，冰上静置 30min，42℃热激 90 s，冰上静置 5 min，37 ℃活化并涂布于平板。

2.蛋白的表达纯化

（1）挑菌摇菌：37 ℃小量（5 ml LB）过夜培养。将过夜培养的菌液接入 800 mlLB 中 37℃水平摇床，220 rpm/min 摇至 OD 值 0.8-1.0。

（2）诱导：向菌液中加入 800 μL IPTG 至终浓度 1 mM。16 ℃诱导 15-20 h。

（3）收菌：将菌液 4000 rpm 离心 15 min，弃掉上清，使用移液器加入适量缓冲液 A（20 mM HEPES PH 7.5，300 mM NaCl），涡旋振荡重悬后可用液氮速冻，并可保存于-80℃冰箱。

（4）超声破碎：加入 1 mM PMSF 于重悬液中，置于冰水混合物中进行超声（220W，20 min，超 3 s，停 5 s）

（5）离心：12000 rpm，4 ℃离心 30 min。

（6）亲和纯化：将 2 ml Cobalt beads 加入进干净的重力柱中，分别用双蒸馏水（ddH_2O）和缓冲液 A 进行平衡。将离心上清加入到重力柱内 3 次，保证样品充分与 beads 结合。然后分别使用缓冲液 A、低咪唑洗脱液（10mM

咪唑，20 mM HEPES PH 7.5，300 mM NaCl）洗脱杂蛋白，最后分批次使用高咪唑洗脱目的蛋白。（200 mM 咪唑，20 mM HEPES PH 7.5，300 mM NaCl）

（7）跑胶：以上各过程均取微量样品制样用于聚丙烯酰胺凝胶电泳（SDS-PAGE）

3.凝胶过滤层析

（1）装柱平衡：将预装柱 Superdex 200 安装到 AKTA 层析系统上。本研究蛋白纯化均用 Superdex 200 预装柱。使用缓冲液 B（20 mM HEPES PH 7.5，150 mM NaCl），以 0.5 ml/min 的流速，平衡 1-2 个柱体积。

（2）上样：选择合适的上样环，首先用缓冲液 B 清洗 3 次，将高浓度目的蛋白样品注入上样环内。

（3）收样：将收样样品管安置好等待出峰收集样品，收集好的样品各取微量进行 SDS-PAGE 确定是否为目的蛋白。

（4）清洗：洗脱结束后，使用缓冲液 B 再清洗 1-2 个柱体积。

2.8 结构预测

使用 AlphaFold (Jumper et al., 2021)对ΦHM7 的 gp40 蛋白进行预测，并利用 Dali 服务器 (http://ekhidna2.biocenter.helsinki.fi/dali/)进行结构比对。

第 3 章 实验结果

3.1 温和型噬菌体ΦHM7 的分离鉴定

实验室前期招募志愿者进行新鲜粪便样品的收集，并使用多种培养方法后通过平板划线分离得到肠道菌的单克隆。对分离到的菌株进行 16s 测序鉴定。选取其中溶木聚糖拟杆菌（*Bacteroides xylanisolvens*）HM7 进行进一步噬菌体分离工作。

我们选用北京市内多处化粪池污水样品及粪便滤液进行噬菌体分离。在经过多次双层平板实验及液体富集实验后，溶木聚糖拟杆菌（Bacteroides xylanisolvens）没有形成明显的噬菌斑，但菌可被部分裂解。因此接下来我们将富集液浓缩制备负染样品，通过透射电子显微镜观察到一种长尾噬菌体，将其命名为ΦHM7。ΦHM7 属于长尾噬菌体科（*Siphoviridae*），其头部衣壳大小约为 60 nm，尾管长度约为 200 nm。

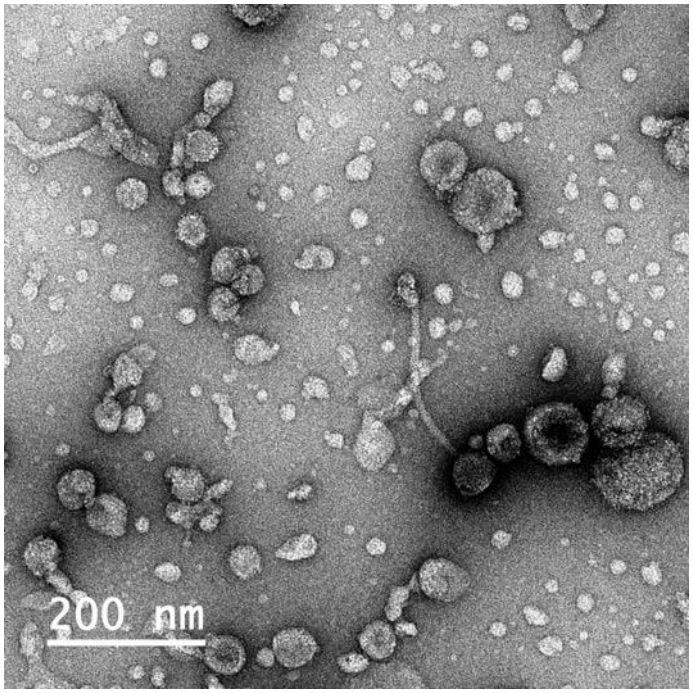

图 3.1 噬菌体ΦHM7 的形态

3.2 ΦHM7 的基因组分析

为了进一步探究噬菌体 HM7 的遗传背景，我们提取了基因组并委托公司进行全基因组测序。HM7 的基因组为 dsDNA，大小为 42,990 bp，GC 含量为 52.4%，对其开放阅读框（ORFs）进行预测显示ΦHM7 编码 45 个蛋白（其中 8 个反向 ORF），此外利用 tRNAscan-SE v. 2.0 (Chan and Lowe, 2019) 工具检测出一个 tRNA Tyr。

我们首先通过 Blastn 在 NT（Nucleotide sequence database）的库中进行核酸序列比对。结果显示与噬菌体 p00（BK010646.1）基因组覆盖率高达 97%，基因相似度 97.2%。此外在 *Bacteroides zhangwenhongii*，*Bacteroides finegoldii*，*Bacteroides fragilis*，*Bacteroides thetaiotaomicron*，*Bacteroides faecis*，*Bacteroides ovatus*，*Bacteroides xylanisolvens*，*Bacteroides uniformis*，*Phocaeicola dorei*，*Bacteroides vulgatus*，*Bacteroides salyersiae*，*Parabacteroides distasonis* 等多个细菌种属中均出现了与 HM7 基因组相似的基因序列，由此可知噬菌体 HM7 是温和型噬菌体，存存在于多种细菌基因组中，具有广泛的宿主。

我们通过 Blastp 对ΦHM7 编码的蛋白功能进行了注释，并确定了这些基因在基因组上的分布情况，通过分析蛋白比对的结果，我们发现多数蛋白显示未知功能，这可能是由于肠道噬菌体目前还未被很好的分离及注释，数据库存在一定的空白，注释结果如图 3.2 所示。

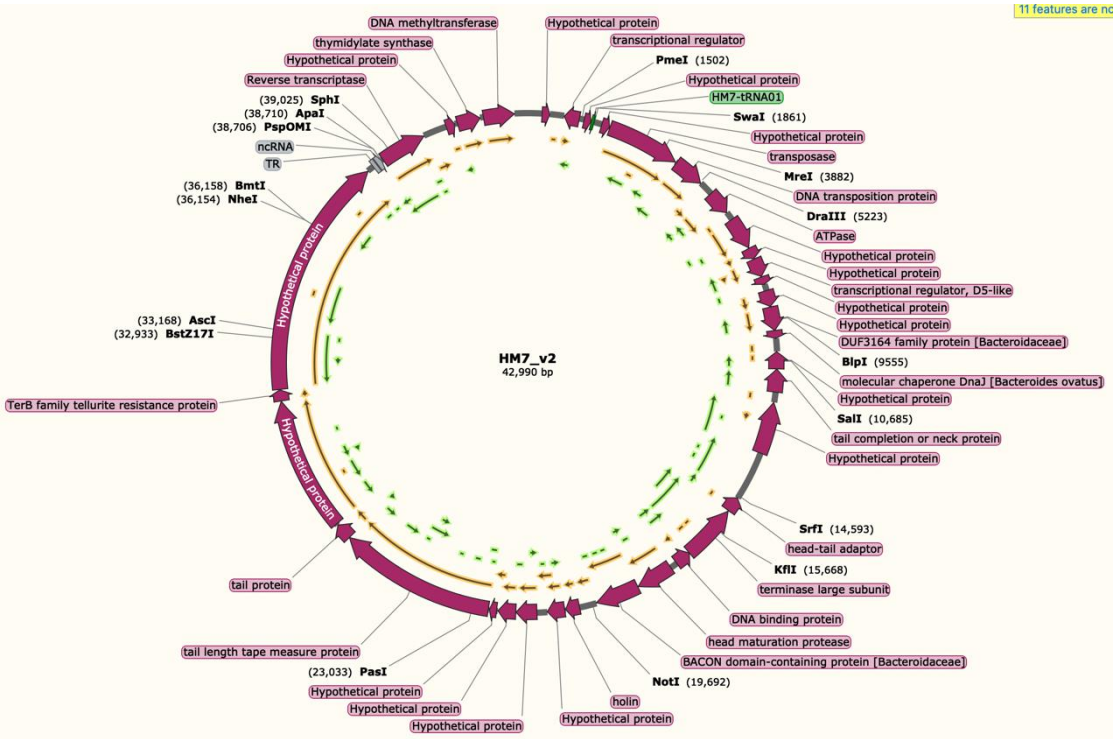

图 3.2 噬菌体ΦHM7 的基因组注释图谱

3.3 ΦHM7 的体外诱导纯化

基于对ΦHM7 基因组的分析，可知它是温和型噬菌体，因此我们想分析其宿主菌基因组中是否整合了ΦHM7 相似的基因序列。我们对 *Bacteroides xylanisolvens HM7* 进行了全基因组测序，并将噬菌体ΦHM7 基因组与细菌

HM7 基因组序列进行比对，发现细菌 HM7 中存在噬菌体ΦHM7 基因组相似的序列，说明我们分离到的细菌 HM7 包含温和型噬菌体ΦHM7。

接下来我们尝试利用细菌 HM7 对温和型噬菌体ΦHM7 进行体外诱导纯化。我们分别选择丝裂霉素，氨苄，卡巴多抗生素进行诱导，结合负染样品的观察，我们发现，氨苄和卡巴多可以诱导出相对完整的噬菌体，而丝裂霉素产生的噬菌体相对不完整，如图 3.3 所示。

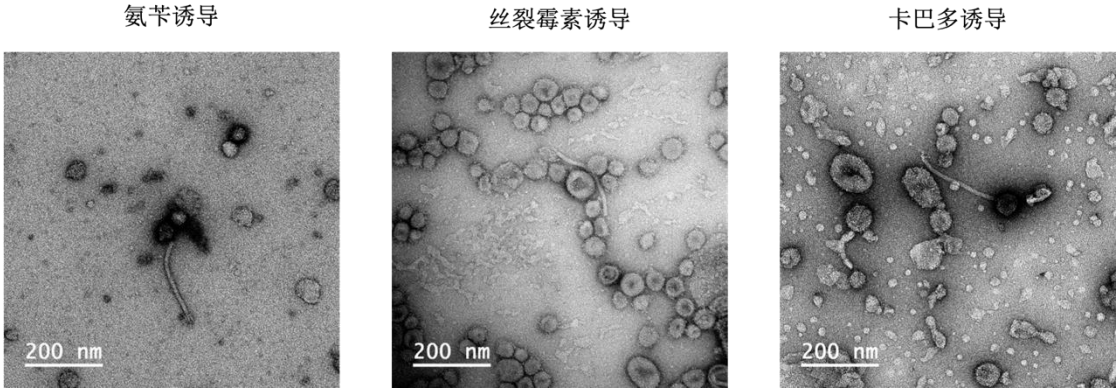

图 3.3 不同种类抗生素诱导ΦHM7 的电镜照片

3.3 ΦHM7 的进化关系分析

为了分析ΦHM7 与其他温和型噬菌体的进化关系，我们选取了之前 blastn 比对得到的前 38 条相似序列，寻找到其中噬菌体末端酶大亚基序列，进行了多序列比对和进化树分析。如图 3.4 所示，我们发现ΦHM7 类似噬菌体的末端酶大亚基序列很保守，相似度在 80%以上。比较噬菌体来源的宿主，可以观察到同种细菌所含的溶源噬菌体序列整体上更相似，ΦHM7 和其他溶木聚糖拟杆菌噬菌体聚成一簇。但是也有部分噬菌体出现在不同种拟杆菌属中，包括福西亚菌属、副拟杆菌属，进一步说明了ΦHM7 具有广泛的宿主范围。从进化关系上看，ΦHM7 和 p00 存在一定进化距离，分属于不同类群，ΦHM7 可以认为是全新的一株噬菌体。

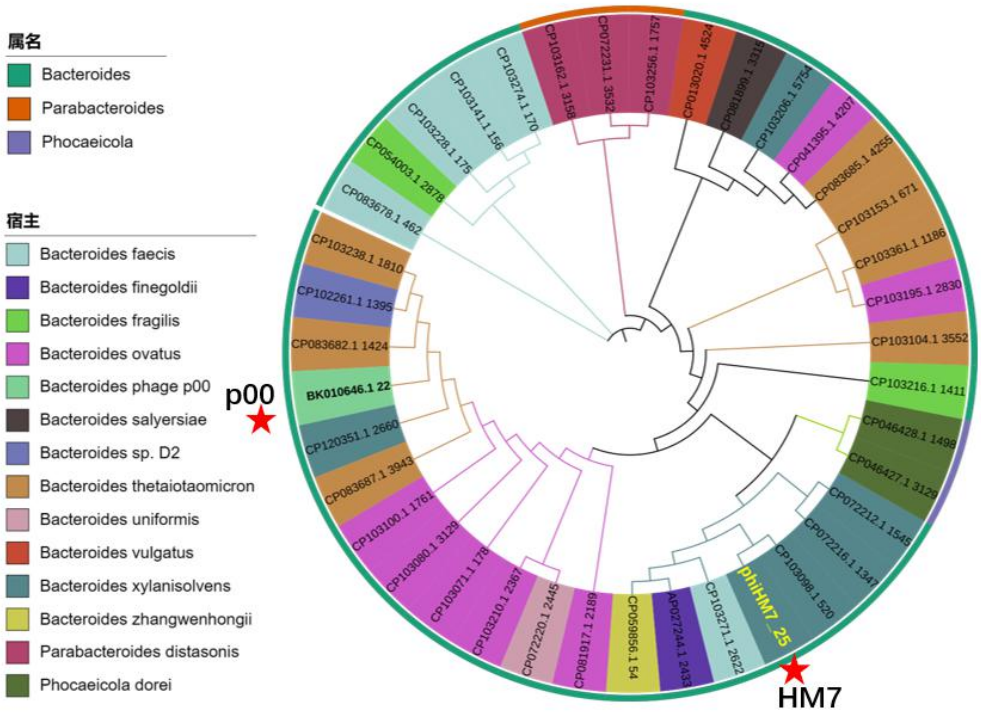

图 3.4 噬菌体ΦHM7 的进化关系分析

图中展示了噬菌体ΦHM7 与噬菌体 p00 的进化关系分析，多种细菌内整合的温和型噬菌体序列与ΦHM7 相似。

通过以上进化关系分析，我们发现存在很多与ΦHM7 同源的噬菌体，这说明ΦHM7 可能广泛存在于人类肠道微生物组中。于是我们利用人类统一肠道病毒组目录 UHGV 数据库对ΦHM7 类似噬菌体进行查询来了解其分布情况。我们发现ΦHM7 在 2349 个宏基因组样本中都能被检测到，远多于常见的 lambda 噬菌体以及丰度最高的裂解型噬菌体 crassphage，说明ΦHM7 在不同人群中广泛存在，如图 3.5 所示。考虑到ΦHM7 是温和型噬菌体，宏基因组数据不能真实反映出病毒颗粒的含量，我们又利用了宏病毒组的数据进行分析。结果显示在 18 个研究共计 189 个样本中都发现了ΦHM7 噬菌体序列，占比 26% (189/729)，如图 3.6 所示。我们发现ΦHM7 噬菌体颗粒广泛存在于婴儿、正常成人肠道内。此外，我们还关注到ΦHM7 与肠炎、肥胖、酒精性脂肪

肝、食管癌等疾病存在密切关系，ΦHM7 可能参与调节人体代谢稳态，具体的机制还有待深入探究。

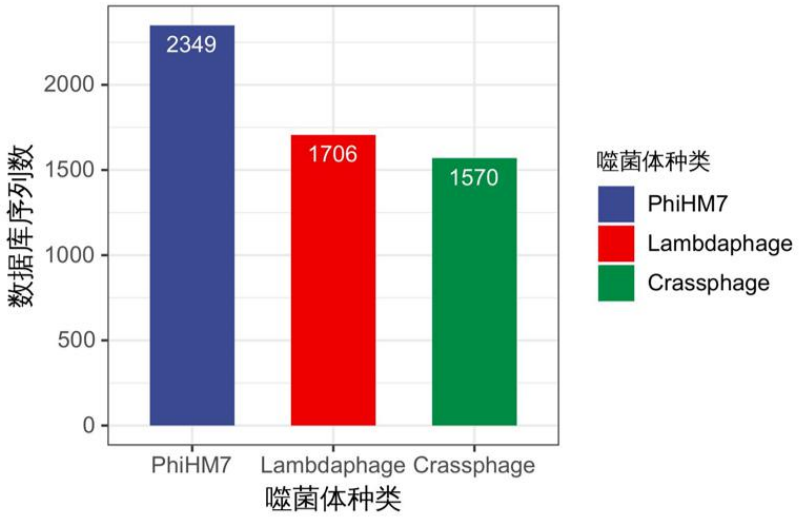

图 3.5 噬菌体ΦHM7 的宏基因组丰度分析

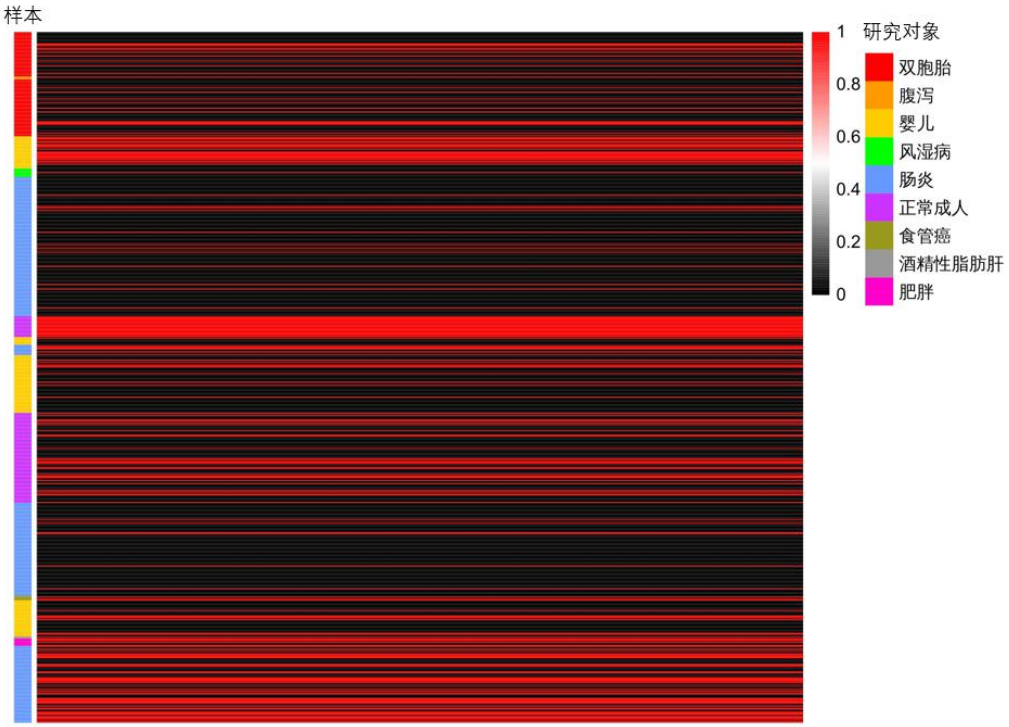

图 3.6 噬菌体ΦHM7 的宏病毒组分析

3.4 DGR 系统可变蛋白 gp40 的鉴定与纯化

我们通过 Blastp 对ΦHM7 编码的蛋白进行分析，发现其存在 DGR 系统。其广泛宿主性可能与其存在 DGR 系统有关，因此我们对其中的 DGR 系统进行重点研究。DGR 系统介导携带保守区序列（TR）的逆转录酶 gp41 靶向识别

可变区序列（VR），其 VR 区存在于 gp40 上，gp40 编码 2240 个氨基酸，Blastp 比对结果为未知功能。由于 DGR 系统的特殊性使得 VR 区可以产生多种突变，从而使噬菌体可以识别更多的细菌靶点，获得更广泛的宿主侵染能力。此前的研究表明 DGR 系统可变区（VR）所在的蛋白一般存在于噬菌体尾部结构，包括尾纤维蛋白、尾刺蛋白、基板蛋白等 (Liu et al., 2002; McMahon et al., 2005; Dai et al., 2010)。利用 AlphaFold 对ΦHM7 的 gp40 进行结构预测，如图 3.7，gp40 的结构包含 5 个结构域，利用 Dali 服务器对每个结构域进行结构相似性比对，结果显示 C 端的结构域 5 与其他噬菌体 DGR 系统靶向蛋白相似，结构域 4 与一些噬菌体的尾纤维蛋白相似，结构域 2 是典型的 IG-like 折叠，这类折叠方式在最近的研究中表明可能存在多糖结合能力，可能与细菌表面多糖或肠道中粘液的黏蛋白相互作用。根据目前已知的一些噬菌体的双尾纤维蛋白结构，我们推测 gp40 蛋白单独就可以自组装形成双尾纤维蛋白结构，如图 3.7 所示。

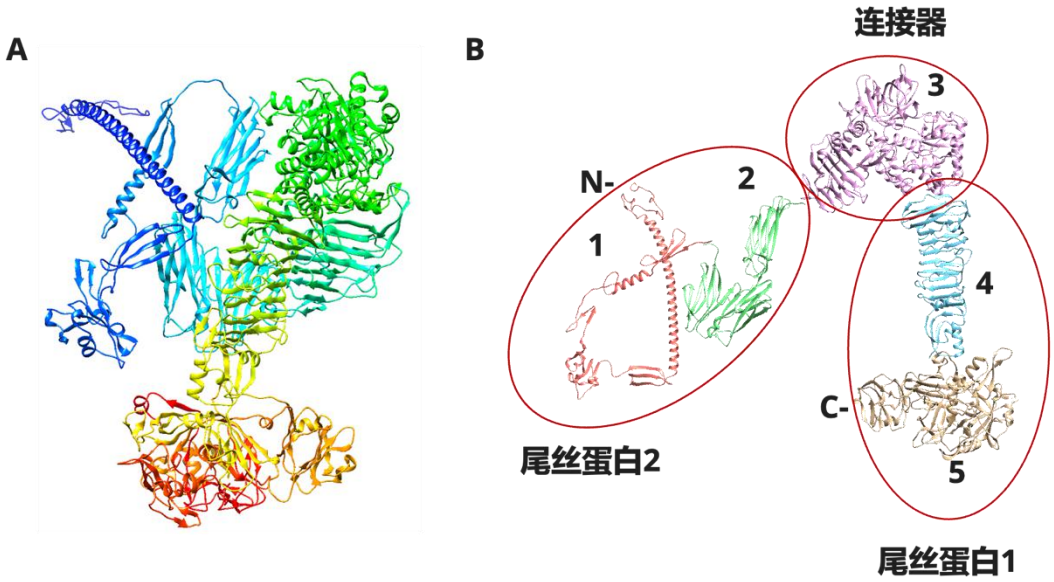

图 3.7 噬菌体ΦHM7 潜在尾丝蛋白 gp40 的预测结构

图 A 为使用 AlphaFold 预测产生的结构。图 B 为基于不同结构域的分析预测。

为了进一步探究 gp40 的功能，利用大肠杆菌对 gp40 蛋白进行体外表达纯化。Gp40 编码 2240 个氨基酸，分子量约为 242.9kDa，将测序正确的质粒转化 BL21 感受态细胞进行诱导表达，IPTG 16℃诱导后将蛋白 NHis-gp40 利用 Cobalt beads 亲和纯化。通过 Western 印迹法确定目的条带为 gp40，如图 3.8。洗脱样品经过 Superdex 200 凝胶过滤层析柱进一步纯化。凝胶过滤层析结果显示 gp40 蛋白有 4 个峰且聚丙烯酰胺凝胶电泳（SDS-PAGE）结果显示第 1、2、3 峰主要为目的蛋白，如图 3.8 所示。

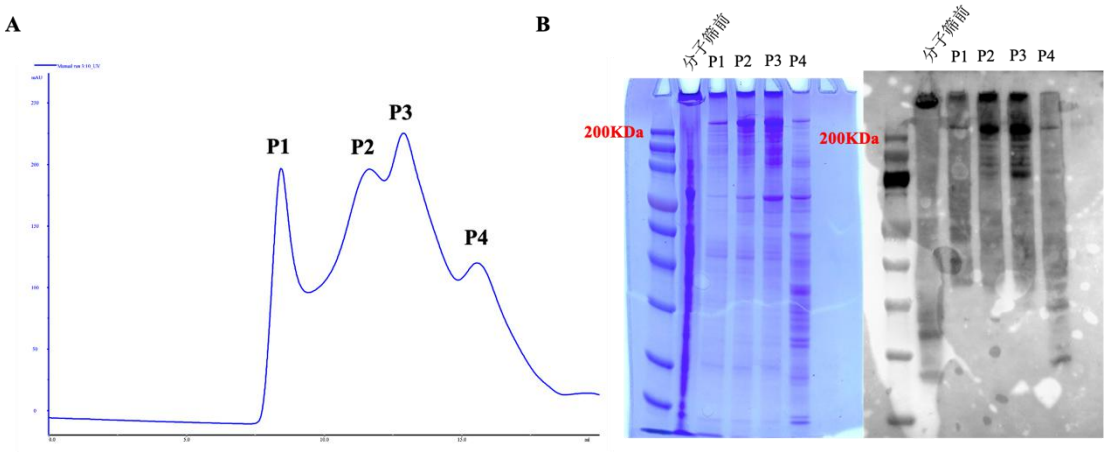

图 3.8 gp40 蛋白的表达纯化

图 A 为分子筛峰形图，包含 4 个峰。图 B 为聚丙烯酰胺凝胶电泳（SDS-PAGE）结果，左侧为考马斯亮蓝染色，右侧为 Western 印迹法。

第 4 章 讨论

在复杂的肠道微生物体系中，除细菌之外还存在多种微生物，如真核病毒、噬菌体、真菌等。其中噬菌体与肠道细菌共同调节肠道微生物的结构与功能，在人类健康与疾病中发挥着重要的功能。但由于肠道内多数噬菌体的获取相对困难，因此目前肠道噬菌体研究还是以宏基因组分析为主，这严重限制了肠道噬菌体的实体研究。本研究通过分离出一株广泛存在于人体肠道内的温和型噬菌体ΦHM7，其具有广泛的宿主侵染能力，这在噬菌体中说比较少见也是非常新颖的。通过宏基因组数据分析发现其广泛存在于人类粪便中，且在健康人群与患病人群中存在一定的差异性，表明ΦHM7 可能潜在影响着人类肠道健康。进一步分析其广泛宿主侵染能力的原因，我们发现ΦHM7 噬菌体存在多样性生成逆转录因子 (DGR)，该系统可以对可变蛋白（例如噬菌体 BPP-1 尾纤维蛋白 Mtd）进行随机突变，使得可变蛋白多样化。多样化的可变蛋白具有不同靶点结合能力，可以靶向到不同的细菌表面受体上实现大范围宿主侵染，这也是ΦHM7 基因组出现在多种拟杆菌细菌基因组上的原因。接下来我们又成功构建了ΦHM7 中 DGR 系统靶向的可变蛋白 gp40 体外纯化的方法，可以实现 gp40 的体外表达纯化，这为之后对于可变蛋白 gp40 功能的研究提供了基础。通过对可变蛋白 gp40 结构进行预测发现其结构非常新颖，可能作为全新的潜在尾丝蛋白出现。总的来说，我们分离到一株具有广泛宿主的拟杆菌噬菌体ΦHM7，其含量在人体肠道内非常丰富，进一步研究发现其广泛宿主性是由其特有的 DGR 系统指导。在之后的研究中我们将重点关注可变蛋白 gp40 是如何发挥功能的，也将通过进一步的结构解析分析其功能的结构机理。该研究将为目前温和型肠道噬菌体的实体研究及 DGR 系统所靶向的潜在尾丝蛋白的功能提供新的思路。

参考文献

Adams, M.J., Lefkowitz, E.J., King, A.M., Harrach, B., Harrison, R.L., Knowles, N.J., et al. (2017). 50 years of the International Committee on Taxonomy of Viruses: progress and prospects. *Arch Virol* 162(5)**,** 1441-1446. doi: 10.1007/s00705-016-3215-y.

Aggarwala, V., Liang, G., and Bushman, F.D. (2017). Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob DNA* 8**,** 12. doi: 10.1186/s13100-017-0095-y.

Benler, S., Cobián-Güemes, A.G., McNair, K., Hung, S.H., Levi, K., Edwards, R., et al. (2018). A diversity-generating retroelement encoded by a globally ubiquitous Bacteroides phage. *Microbiome* 6(1)**,** 191. doi: 10.1186/s40168-018-0573-6.

Burokas, A., Moloney, R.D., Dinan, T.G., and Cryan, J.F. (2015). Microbiota regulation of the Mammalian gut-brain axis. *Adv Appl Microbiol* 91**,** 1-62. doi: 10.1016/bs.aambs.2015.02.001.

Camarillo-Guerrero, L.F., Almeida, A., Rangel-Pineros, G., Finn, R.D., and Lawley, T.D. (2021). Massive expansion of human gut bacteriophage diversity. *Cell* 184(4)**,** 1098-1109.e1099. doi: 10.1016/j.cell.2021.01.029.

Chan, P.P., and Lowe, T.M. (2019). tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol* 1962**,** 1-14. doi: 10.1007/978-1-4939-9173-0_1.

Couto, M.R., Gonçalves, P., Magro, F., and Martel, F. (2020). Microbiota-derived butyrate regulates intestinal inflammation: Focus on inflammatory bowel disease. *Pharmacol Res* 159**,** 104947. doi: 10.1016/j.phrs.2020.104947.

Dai, W., Hodes, A., Hui, W.H., Gingery, M., Miller, J.F., and Zhou, Z.H. (2010). Three-dimensional structure of tropism-switching Bordetella bacteriophage. *Proc Natl Acad Sci U S A* 107(9)**,** 4347-4352. doi: 10.1073/pnas.0915008107.

Davidson, A.R., Cardarelli, L., Pell, L.G., Radford, D.R., and Maxwell, K.L. (2012). Long noncontractile tail machines of bacteriophages. *Adv Exp Med Biol* 726**,** 115-142. doi: 10.1007/978-1-4614-0980-9_6.

Dion, M.B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nat Rev Microbiol* 18(3)**,** 125-138. doi: 10.1038/s41579-019-0311-5.

Doulatov, S., Hodes, A., Dai, L., Mandhana, N., Liu, M., Deora, R., et al. (2004). Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. *Nature* 431(7007)**,** 476-481. doi: 10.1038/nature02833.

Ghosh, T.S., Shanahan, F., and O'Toole, P.W. (2022). The gut microbiome as a modulator of healthy ageing. *Nat Rev Gastroenterol Hepatol* 19(9)**,** 565-584. doi: 10.1038/s41575-022-00605-x.

Guo, H., Tse, L.V., Barbalat, R., Sivaamnuaiphorn, S., Xu, M., Doulatov, S., et al. (2008). Diversity-generating retroelement homing regenerates target sequences for repeated rounds of codon rewriting and protein diversification. *Mol Cell* 31(6)**,** 813-823. doi: 10.1016/j.molcel.2008.07.022.

Hohn, T., and Katsura, I. (1977). Structure and assembly of bacteriophage lambda. *Curr Top Microbiol Immunol* 78**,** 69-110. doi: 10.1007/978-3-642-66800-5_3.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873)**,** 583-589. doi: 10.1038/s41586-021-03819-2.

Lebeer, S., Bron, P.A., Marco, M.L., Van Pijkeren, J.P., O'Connell Motherway, M., Hill, C., et al. (2018). Identification of probiotic effector molecules: present state and future perspectives. *Curr Opin Biotechnol* 49**,** 217-223. doi: 10.1016/j.copbio.2017.10.007.

Liu, M., Deora, R., Doulatov, S.R., Gingery, M., Eiserling, F.A., Preston, A., et al. (2002). Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. *Science* 295(5562)**,** 2091-2094. doi: 10.1126/science.1067467.

Lobb, B., Tremblay, B.J., Moreno-Hagelsieb, G., and Doxey, A.C. (2020). An assessment of genome annotation coverage across the bacterial tree of life. *Microb Genom* 6(3). doi: 10.1099/mgen.0.000341.

Macadangdang, B.R., Makanani, S.K., and Miller, J.F. (2022). Accelerated Evolution by Diversity-Generating Retroelements. *Annu Rev Microbiol* 76**,** 389-411. doi: 10.1146/annurev-micro-030322-040423.

Marvin, D.A., Symmons, M.F., and Straus, S.K. (2014). Structure and assembly of filamentous bacteriophages. *Prog Biophys Mol Biol* 114(2)**,** 80-122. doi: 10.1016/j.pbiomolbio.2014.02.003.

McMahon, S.A., Miller, J.L., Lawton, J.A., Kerkow, D.E., Hodes, A., Marti-Renom, M.A., et al. (2005). The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat Struct Mol Biol* 12(10)**,** 886-892. doi: 10.1038/nsmb992.

Miller, J.L., Le Coq, J., Hodes, A., Barbalat, R., Miller, J.F., and Ghosh, P. (2008). Selective ligand recognition by a diversity-generating retroelement variable protein. *PLoS Biol* 6(6)**,** e131. doi: 10.1371/journal.pbio.0060131.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., et al. (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 21(10)**,** 1616-1625. doi: 10.1101/gr.122705.111.

Ott, S.J., Waetzig, G.H., Rehman, A., Moltzau-Anderson, J., Bharti, R., Grasis, J.A., et al. (2017). Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With Clostridium difficile Infection. *Gastroenterology* 152(4)**,** 799-811.e797. doi: 10.1053/j.gastro.2016.11.010.

Paul, B.G., Burstein, D., Castelle, C.J., Handa, S., Arambula, D., Czornyj, E., et al. (2017). Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. *Nat Microbiol* 2**,** 17045. doi: 10.1038/nmicrobiol.2017.45.

Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., et al. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466(7304)**,** 334-338. doi: 10.1038/nature09199.

Roux, S., Paul, B.G., Bagby, S.C., Nayfach, S., Allen, M.A., Attwood, G., et al. (2021). Ecology and molecular targets of hypermutation in the global microbiome. *Nat Commun* 12(1)**,** 3076. doi: 10.1038/s41467-021-23402-7.

Schiavi, E., Gleinser, M., Molloy, E., Groeger, D., Frei, R., Ferstl, R., et al. (2016). The Surface-Associated Exopolysaccharide of Bifidobacterium longum 35624 Plays an Essential Role in Dampening Host Proinflammatory Responses and Repressing Local TH17 Responses. *Appl Environ Microbiol* 82(24)**,** 7185-7196. doi: 10.1128/aem.02238-16.

Shkoporov, A.N., and Hill, C. (2019). Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. *Cell Host Microbe* 25(2)**,** 195-209. doi: 10.1016/j.chom.2019.01.017.

Sun, L., Young, L.N., Zhang, X., Boudko, S.P., Fokine, A., Zbornik, E., et al. (2014). Icosahedral bacteriophage ΦX174 forms a tail for DNA transport during infection. *Nature* 505(7483)**,** 432-435. doi: 10.1038/nature12816.

致谢

肠道噬菌体作为肠道内最丰富的微生物在人类健康与疾病中发挥着至关重要的作用。但由于肠道内多数噬菌体的获取相对困难，因此目前肠道噬菌体研究还是以宏基因组分析为主，这严重限制了肠道噬菌体的实体研究。本人在前期自主学习以及相关肠道菌群、肠道噬菌体分离实验的基础上，聚焦肠道温和型噬菌体的分离鉴定，并基于其具有广泛的宿主侵染能力，在测序后发现其含有DGR系统后确定了进一步的研究目标。

作为本小组唯一的学生组员，本人主要负责开展噬菌体分离鉴定的相关实验、噬菌体序列注释和研究分析，以及论文的撰写等工作。

在开展本研究的过程中，我的“英才计划”导师——清华大学向烨教授、刘哲博士在选题、理论、实验以及论文写作等方面无偿给予了诸多建设性的指导和帮助，既进一步培养了我的科研兴趣，极大地提升了我的科研能力，也为本研究的顺利推进奠定了坚实的基础。

在这里，向一直以来辛苦指导、精心培养我的导师和学长们表达我最真诚的感谢和最崇高的敬意！也向一直关心、鼓励我完成本研究的爸爸妈妈表示最衷心的谢意！正是因为有你们的鼎力支持，我将在生物科研的这条大道上继续奋勇前行，为不断提升人类健康水平而不懈奋斗！

【团队成员和指导老师情况】

参赛队员：王兆琪 北京师范大学附属实验中学国际部高三学生（2024届）

指导老师：向 烨

简历：向烨研究员1998年本科毕业于北京大学化学系，2003年获中科院生物物理所博士学位。2005年1月赴美国普渡大学从事博士后研究，并于2010年晋升为助理研究科学家。2012年入选中组部青年计划。2019年入选国家杰青。2013年1月回清华大学任教，现为清华大学医学院研究员，博士生导师，长聘副教授。向烨研究员的科研重点是人类重大病毒性传染病的致病机理，病毒与免疫系统相互作用关系，抗病毒药物和疫苗研发。主要研究方向包括乙肝病毒，登革热病毒，甲病毒、新型冠状病毒肺炎病毒等高致病性病毒。近5年以在Nature，Science等杂志发表20多篇原创性论文，论文被引3200多次。研究建立了病毒入侵宿主细胞分子机器详细机制，为相应疫苗药物研究提供理论指导。

指导老师：刘 哲 清华大学基础医学博士生