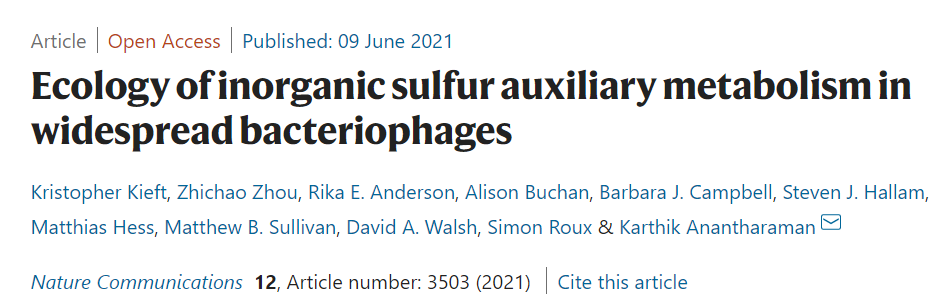
**Nature Communications | 噬菌体中无机硫辅助代谢基因的生态学研究**

翻译：周之超@UW-Madison

**Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages** ****

Article

Nature Commnications, [IF 12.121]

DOI：10.1038/s41467-021-23698-5

原文链接：https://www.nature.com/articles/s41467-021-23698-5

第一作者：Kristopher Kieft, Zhichao Zhou（周之超）

通讯作者：Karthik Anantharaman

主要单位：Department of Bacteriology, University of Wisconsin - Madison威斯康星大学麦迪逊分校细菌系

**摘要**

**Abstract**

微生物的硫代谢有助于促进全球范围内的生物地球化学循环。硫代谢微生物被噬菌体感染，这些噬菌体可以编码辅助代谢基因（AMG）来改变宿主细胞内的硫代谢，但其具体生态学功能研究，现阶段仍然很有限。在此次研究中，我们确定了191个噬菌体，它们来自12个环境，编码了227个具有硫和硫代硫酸盐氧化功能的AMG（dsrA、dsrC/tusE、soxC、soxD和soxYZ）。在不同噬菌体种群的生态位分化过程中保留AMG的证据表明，辅助代谢赋予噬菌体可被测量的适应度提升，并对生态系统中生物地化循环产生重要影响。AMG基因丰度和表达谱表明，噬菌体对淡水湖泊和海洋中的硫和硫代硫酸盐的氧化有显著贡献，以及对热液环境中硫浓度变化有着敏感反应。总的来说，我们提供了关于与硫相关的噬菌体辅助代谢的分布、多样性和生态学的基础性研究，并重申了将病毒的贡献纳入生物地球化学框架中的必要性。

**引言**

**Introduction**

感染细菌的病毒（噬菌体）被认为会比其细菌宿主有更大的基因功能库，并且在整个微生物群落中转移基因方面非常活跃。大多数已知的噬菌体都通过最小化非编码区域、减少编码蛋白质的平均长度、融合蛋白质并保留少数非必要基因来进化出紧凑的基因组。尽管噬菌体的基因组大小较小，编码能力有限，但它们以在感染过程中调节宿主细胞、接管细胞代谢过程并通过细菌增殖而闻名，这些活动通常是通过宿主细胞的裂解。噬菌体感染的宿主，被称为病毒感染细胞（virocell），与未感染的状态相比，具有独特的生理学特征。根据一些估计，在水生环境中，多达20-40%的细菌被认为处于病毒感染细胞状态，正在经历噬菌体导向的新陈代谢。这已经引起了人们对理解噬菌体在宿主内得到营养代谢重定向能力的机制，以及这种操作最终如何影响微生物体和生态系统的浓厚兴趣。

噬菌体改变其宿主代谢状态的其中一种机制是通过噬菌体编码的辅助代谢基因（AMG）的活性来改变其代谢状态。AMG通常从宿主细胞获得（即重组到噬菌体基因组上），可在感染期间用来增强或重定向宿主细胞内的特异性代谢过程。这些增强功能可能可以维持、驱动或短路化代谢途径的重要步骤，并可以为噬菌体在特定的代谢或营养条件下提供足够的适合度优势，以便随着时间的推移保留这些基因。AMG的两个显著例子是核心光系统II蛋白的编码基因*psbA*和*psbD*，它们通常由在淡水和海洋环境中感染蓝藻的噬菌体编码，并负责在感染期间补充病毒感染细胞的光系统功能。PsbA和PsbD在维持宿主内光合能量的产生方面起着重要作用；这种能量随后被用于生产噬菌体传播的资源（如核苷酸）。蓝藻宿主不受益于额外的基因拷贝（即噬菌体AMG），因为复制的好处和能量获取有利于感染的噬菌体。对AMG的其他描述包括远洋海洋中的硫氧化、淡水湖中的甲烷氧化、地表海洋中的氨氧化、土壤中的碳利用（如碳水化合物水解）和海洋氨化。作为进一步的一个例子，有研究者假设，一些编码碳利用AMG的噬菌体通过诱导宿主进入饥饿状态，将碳从糖酵解重定向到dNTP合成途径，以进行噬菌体基因组的复制。在这种情况下，噬菌体编码他们自己的AMG来特定地操作宿主过程，而不是简单地向宿主提供一个额外的基因拷贝。除了这些例子之外，噬菌体辅助代谢在生态系统尺度水平水平的综合影响尚未被充分探索，且在微生物群落功能和相互作用的背景下也未被具体描述。

异化硫代谢（DSM）包括还原（如硫酸盐转化为硫化物）和氧化（如硫化物或硫代硫酸盐转化为硫酸盐），它们在地球上硫代谢中占有较重地位。具有DSM功能的细菌（称为硫微生物）在系统发育上是多样的，它们可以跨越多个独立的门，可以在一系列自然和人类系统、水生和陆地生物群落、有氧或厌氧环境以及在有光或无光环境中被发现存在。因为DSM通常与初级生产和埋藏的有机碳的转化相关联，理解这些过程对于解释微生物和噬菌体介导的营养物质和能量转化的生物地球化学意义至关重要。感染DSM微生物的噬菌体的研究现在还很有限，仅见于感染γ-变形菌门的SUP05硫细菌、编码*dsrA*和*dsrC*基因的噬菌体和感染表层海洋中变形菌门细菌、编码*dsrC*和*soxYZ*基因的噬菌体。尽管在多个宿主群和环境中发现了DSM AMG，但我们对其全球多样性和在硫的生物地化循环中的作用仍然所知有限。研究描述与DSM相关的噬菌体的生态、功能和作用对于整体理解硫系物质转化和代谢机制至关重要。

在这里，我们利用公开的宏基因组和宏转录组数据来识别能够在宿主细胞内操纵DSM的噬菌体。我们在沿海、远洋、热液喷口、人类和陆地环境中，鉴定了191个编码氧化和歧化还原性硫（如元素硫和硫代硫酸盐）AMG的噬菌体。我们将这些编码DSM AMG的噬菌体称为硫噬菌体。这些硫噬菌体来自*Caudovirales*的不同分类分支，其中包括*Siphoviridae*, *Myoviridae* 和*Podoviridae*；它们基因库和进化背景非常多样化。利用对热液环境、淡水湖和*Tara* Ocean海洋样本中恢复的宏基因组的配对病毒-宿主基因丰度的检测，我们提供了病毒AMG对硫和硫代硫酸盐氧化有显著贡献的证据。对宏转录组数据的调查表明，噬菌体导向的硫氧化活性随着热液生态系统中底物供应的增加而显著增加，这表明病毒感染细胞对不断变化的环境条件做出了快速而敏感的反应。总的来说，我们提供了关于噬菌体导向的异化性硫和硫代硫酸盐代谢的分布、多样性和生态学的基础研究，并且再次重申了将病毒贡献纳入到生物地化循环评估中的必要性。

**结果**

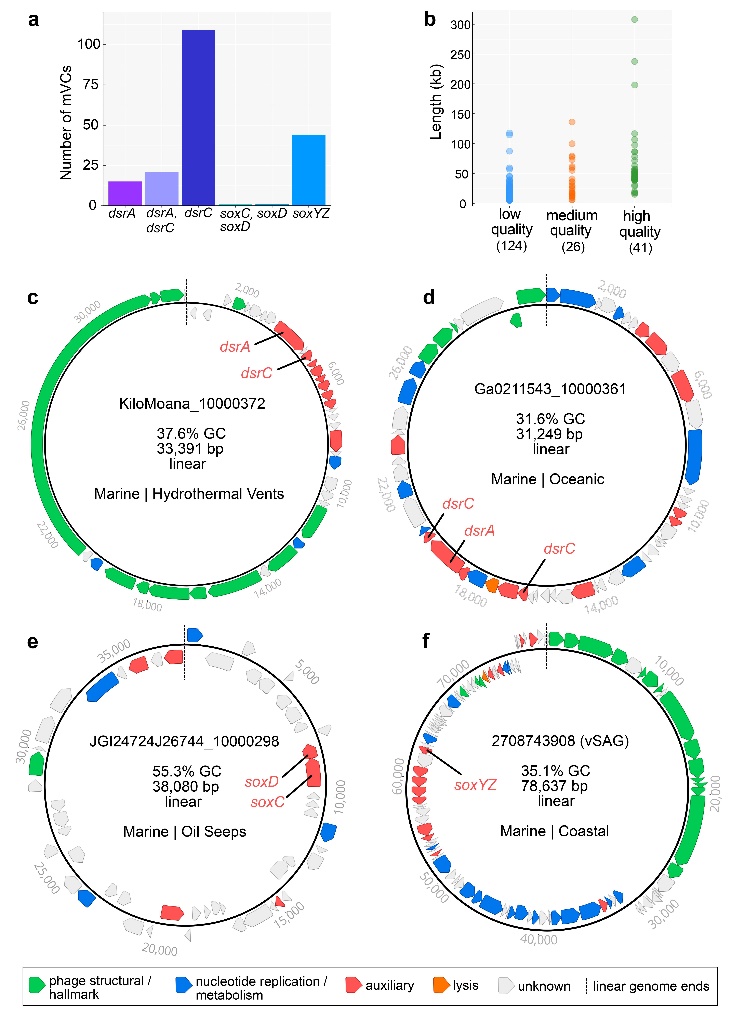
**Result**

**独特的硫噬菌体编码元素硫和硫代硫酸盐氧化AMG**

**Unique sulfur phages encode AMGs for oxidation of elemental sulfur and thiosulfate**

我们查询了Integrated Microbial Genomes/Viruses （IMG/VR v2.1）数据库，以查找编码与异化硫氧化和还原过程相关的噬菌体基因。我们鉴定了190个宏基因组病毒片段（mVC）和一个单扩增病毒基因组，这些基因组编码反向亚硫酸盐还原酶亚基A和C（*dsrA*和*dsrC*）、硫脲合酶亚基E（*tusE*、*dsrC*）、磺烷脱氢酶亚基C和D（*soxC*、*soxD*）以及融合的用于硫代硫酸盐氧化的硫载体蛋白Y和Z (soxYZ)。除一个外（KiloMoana\_10000689），所有mVC估计为部分病毒基因组。虽然携带*dsrA*、*dsrC*/*tusE*和*soxYZ*的噬菌体之前已经在特定的海洋环境中被描述过，但此次是在病毒基因组上编码的*soxC*和*soxD*的首次报告。每个已识别的mVC编码1到4个DSM AMG，共计有227个AMG（图1a）。mVC的长度从5kb到308kb不等，平均长度约为31kb，共有83个序列大于20kb。根据基于基因组的质量估计，所有mVC包含124个低质量、26个中质量和41个高质量草图（图1b）。按照之前研究描述的方法来鉴定，只有一个mVC是一个完整的环形基因组。本研究中的大多数病毒，除了几种编码类*tusE* AMG的mVC外，基于编码的蛋白质功能都被预测有专性裂解宿主的生活方式。

mVC显示出独特和不同的基因组安排特征，无论编码的AMG是什么。但在大多数情况下，编码的AMG是在辅助基因基因盒中发现的，独立于结构基因基因盒和核苷酸代谢基因盒（图1c，d，e，f）。噬菌体中的辅助基因基因盒通常编码的基因不是生产繁殖所必需的，但可以在感染期间提供选择性优势，例如在特定的营养限制条件下或克服代谢瓶颈。这种基因组安排表明，DSM AMG的作用与宿主调节有关，而不是与病毒的基本生活任务相关，如转录/翻译、基因组复制或结构组装。



**图1**数据集汇总统计数据和具有代表性的mVC基因组结构示意图。（a）mVC的数量，总共191个，编码单个或多个DSM AMG。（b）估计的mVC基因组质量比序列长度。mVC编码（c）*dsrA*和*dsrC*、（d）*dsrA*和两个*dsrC*、（e）*soxC*和*soxD*，以及（f）*soxYZ*。对于（c）、（d）、（e）和（f），线性mVC序列可视化为环形，端点用虚线来表示，预测的开放阅读框是根据VIBRANT所注释的功能来着色的。缩写：vSAG代表病毒单扩增基因组。

**AMG蛋白中保守的氨基酸残基和结构域的验证**

**Validation of conserved amino acid residues and domains in AMG proteins**

验证AMG蛋白序列可以确保它们在mVC基因组上的识别代表了准确的注释（即预测的生物学功能）。我们通过生信方法进行蛋白质验证，通过将AMG蛋白序列与分离细菌或噬菌体经过生物化学验证后的参考序列对齐来验证，并评估功能域和保守氨基酸残基的存在与否。我们着重关注辅因子协调/活性位点、细胞色素*c*基序、底物结合基序、siroheme结合位点、半胱氨酸基序和其他严格保守的残基（下面统称为残基）。最后，我们评估了噬菌体AMG是否存在选择压力而被保存。在AMG蛋白序列上鉴定的保守残基包括：DsrA：底物结合（R、KxKxK、R、HeR）和siroheme结合（CxgxxxC、CxxdC）；DsrC：严格保守的半胱氨酸基序（CxxxgxpxpxxxC）；SoxYZ：底物结合半胱氨酸（ggCs）和可变半胱氨酸基序（CC）；SoxC：辅因子协调/活性位点（XxH、D、R、XxK）；SoxD：细胞色素*c*基序（CxxCHG、CMxxC）。对大多数AMG蛋白序列上的这些残基的鉴定表明，它们都具有功能。然而，有几个例子的AMG可能编码非功能或明显不同的基因。例如，只有23个DsrC AMG蛋白序列包含两个严格保守的半胱氨酸基序，112个只包含第二个半胱氨酸基序，1个只包含第一个半胱氨酸基序，另外5个则不包含第一个半胱氨酸基序。噬菌体DsrC中缺乏严格保守的半胱氨酸基序被认为在感染过程中有其他功能，但这一假设尚未得到证实。大多数缺乏一个或多个半胱氨酸残基的DsrC AMG蛋白序列可能在功能上充当TusE，一种用于tRNA硫醇修饰的相关硫转移蛋白。事实上，一些来自人类口腔微生物组的mVC编码了位于其他*tus*基因两侧的类*tusE* AMG。残基缺失的进一步例子包括两个编码SoxD的mVC，其中一个缺失了第一个细胞色素*c*基序，并且都缺少了第二个细胞色素*c*基序。这最初表明了无功能SoxD的存在，但这个想法被SoxC存在保守的残基所质疑。在其中一个mVC中编码的毗邻SoxD的功能SoxC，表明两者都有可能保留功能。研究表明，不同于各自细菌同源物的噬菌体蛋白可以保留其原来的预期活性或提供额外的功能。总的来说，除了118个类*tusE*的AMG外，在对AMG蛋白序列的生信分析中表明，mVC编码有功能的代谢蛋白。

为了了解AMG上的选择性压力，我们计算了噬菌体AMG及其细菌同源基因中非同义与同义核苷酸 比率（dN/dS）的差异，以评估噬菌体基因是否处于纯化（稳定）选择之下。计算出的dN/dS比率低于1表明，一个基因或整个基因组正处于选择压力下，以消除有害突变。因此，通过计算mVC AMG的dN/dS值低于1表明病毒通过消除有害突变选择性地保留AMG的功能。mVC *dsrA*、*dsrC*和*soxYZ* AMG的dN/dS计算结果值低于1，表明AMG处于纯化选择。

**DSM AMG可能操纵硫氧化途径中的关键步骤来重新分配能量**

**DSM AMGs likely manipulate key steps in sulfur oxidation pathways to redistribute energy**

如前所述，mVC编码的DSM AMG可能专门在感染期间操纵宿主细胞中的硫转化。为了更好地理解这种操作的含义，我们构建了未感染和感染宿主中硫（即*dsr* AMG）氧化和硫代硫酸盐(即*sox* AMG)氧化/歧化代谢的概念图（图2）。

为了特别了解携带*dsrC*和*dsrA* AMG的潜在优势，需要考虑硫化物氧化途径中的每一步。在仅限宿主的硫化物氧化期间，扩散到细胞中的硫化物通过硫化物：醌氧化还原酶（如Sqr）转化为元素硫，在某些情况下，该途径可以直接从运入元素硫开始。元素硫可以储存在局部硫球中，直到通过硫化物氧化途径代谢。在硫化物氧化过程中，硫载体蛋白DsrC携带的元素硫被酶复合物DsrAB氧化成亚硫酸盐。这一步被估计是完全路径中的速率限制性步骤，并为ATP产生提供最多的电子（六个电子）。速率限制是由DsrAB酶复合物或DsrC载体的饱和引起的。硫化物/硫氧化的最后步骤包括将亚硫酸盐进一步氧化为5-磷酸硫酸腺苷（APS），然后分别通过APS还原酶（如AprAB）和硫酸盐腺苷转移酶（如Sat）氧化为硫酸盐，从而产生两个电子。然后获得的ATP可以用于细胞过程。相反，在涉及到硫化物氧化调节的噬菌体感染期间，速率限制性步骤（即DsrC和DsrA的协同活性）可被噬菌体DsrC和/或DsrA补充，以潜在地提高反应的速率和ATP产率，以及利用任何存储的元素硫。ATP提供的能量可以有效地用于噬菌体的传播（例如，噬菌体蛋白的产生、基因组复制或基因组封装）（图2a）。

同样地，编码*soxYZ*、*soxC*和*soxD*的噬菌体也可以增强硫代硫酸盐氧化/歧化代谢的正常状态。在仅限宿主的硫代硫酸盐氧化过程中，硫代硫酸盐被运输到细胞中，由SoxYZ运输的两个硫代醇基团经历一系列的氧化反应。部分携带的硫，在产生两个电子后，将以硫酸盐的形式运出细胞。剩余携带的硫可以储存在元素硫球体中，或者进入关键的能量产生步骤。关键的能量产生步骤绕过元素硫的储存，并利用SoxCD酶复合物产生6个电子，使得ATP产生。在涉及硫代硫酸盐氧化/歧化代谢调节的噬菌体感染过程中，宿主和噬菌体SoxYZ硫载体都可以支持整个途径，以持续驱动元素硫储存，然后被Dsr复合物氧化。然而，没有证据表明噬菌体受益于耦合sox和dsr途径，因为没有发现mVC同时编码sox和dsr AMG。最后，可以利用噬菌体SoxCD来驱动途径进入关键的能量产生步骤。与dsr途径一样，由此产生的ATP将被用于噬菌体的传播（图2b）。

Diagram

Description automatically generated

**图2**病毒DsrA、DsrC、SoxC、SoxD和SoxYZ辅助代谢概念图。（a）硫化氢和储存的无机硫的微生物异化性氧化。用于细胞过程和生长的ATP的产生以及通路的速率限制性步骤用星号（左）表示。病毒感染和通过编码的DsrA或DsrC操作硫氧化，以提升途径的速率限制性步骤，并增加用于病毒复制的能量产量（右）。（b）微生物异化性的硫代硫酸盐还原氧化或储存无机硫。由此产生的ATP用于细胞过程和通路的关键能量产生反应，表示为星号（左）。病毒感染和通过编码SoxC、SoxD或SoxYZ氧化硫代硫酸盐，以增强整个途径和关键的能量产生步骤，以增加用于病毒复制的能量产量（右）。对于（a）和（b），细胞过程为红色，硫氧化途径为黑色，能量流为蓝色，病毒过程为橙色（a）或紫色（b）。对于所有的途径步骤，微生物酶和硫载体与病毒增强功能协同作用。

**硫噬菌体广泛分布在环境中**

**Sulfur phages are widely distributed in the environment**

接下来，我们研究了编码DSM AMG的mVC的生态和分布模式。我们通过使用识别的AMG和参考微生物蛋白构建系统发育树，并解析IMG/VR数据库中mVC数据的环境信息，来描述它们在不同环境中不同的生态和分布模式。我们发现编码*dsrA*的mVC主要分布在一些海洋环境中，而编码*dsrC*的mVC则分布在海洋、咸水、石油渗透地带、陆地、工程和共生环境中（图3a，b）。对于*soxC*和*soxD*，我们只识别到了编码这些AMG 的两个宏基因组数据集，一个来自圣巴巴拉通道的石油渗漏地带（编码*soxC*和*soxD*），另一个来自华盛顿湖的淡水沉积物（图3c，d）。编码*soxYZ*的mVC是在水生环境中发现的，包括不同的海洋、咸水和淡水生态系统类型（图3e）。除了在不同生态系统类型中的mVC分布外，我们还确定了它们在全球水平上广泛的生物地理分布（图3f）。总的来说，这些DSM AMG有着广泛的生态地理分布，并可能在许多不同的环境类型和营养条件（包括自然和工程环境）下辅助宿主的代谢功能。

Diagram, schematic

Description automatically generated

**图3** AMG蛋白的系统发育树和噬菌体基因组的世界地理分布。（a，b）噬菌体DsrA和DsrC （c，d，e）SoxC、SoxD和SoxYZ的系统发育树。Ultrafast bootstrap（UFBoot）支持值（>50%）在节点上进行标记。（c，d）噬菌体基因编码的蛋白质序列用星号标记，其环境条件来源信息也在图上有相应的标记。噬菌体基因组的生态系统类型（内环）和生态系统类别（外环）在a，b，e系统发育树中有标记；不同的颜色代表不同的生态系统类型和类别，每个环中的空白位置为参考微生物蛋白。（f）显示含有硫相关AMG的噬菌体基因组分布的世界地图。该地图中人类系统的有关研究被排除在外。不同的生态系统类型由不同的符号和颜色来表示。

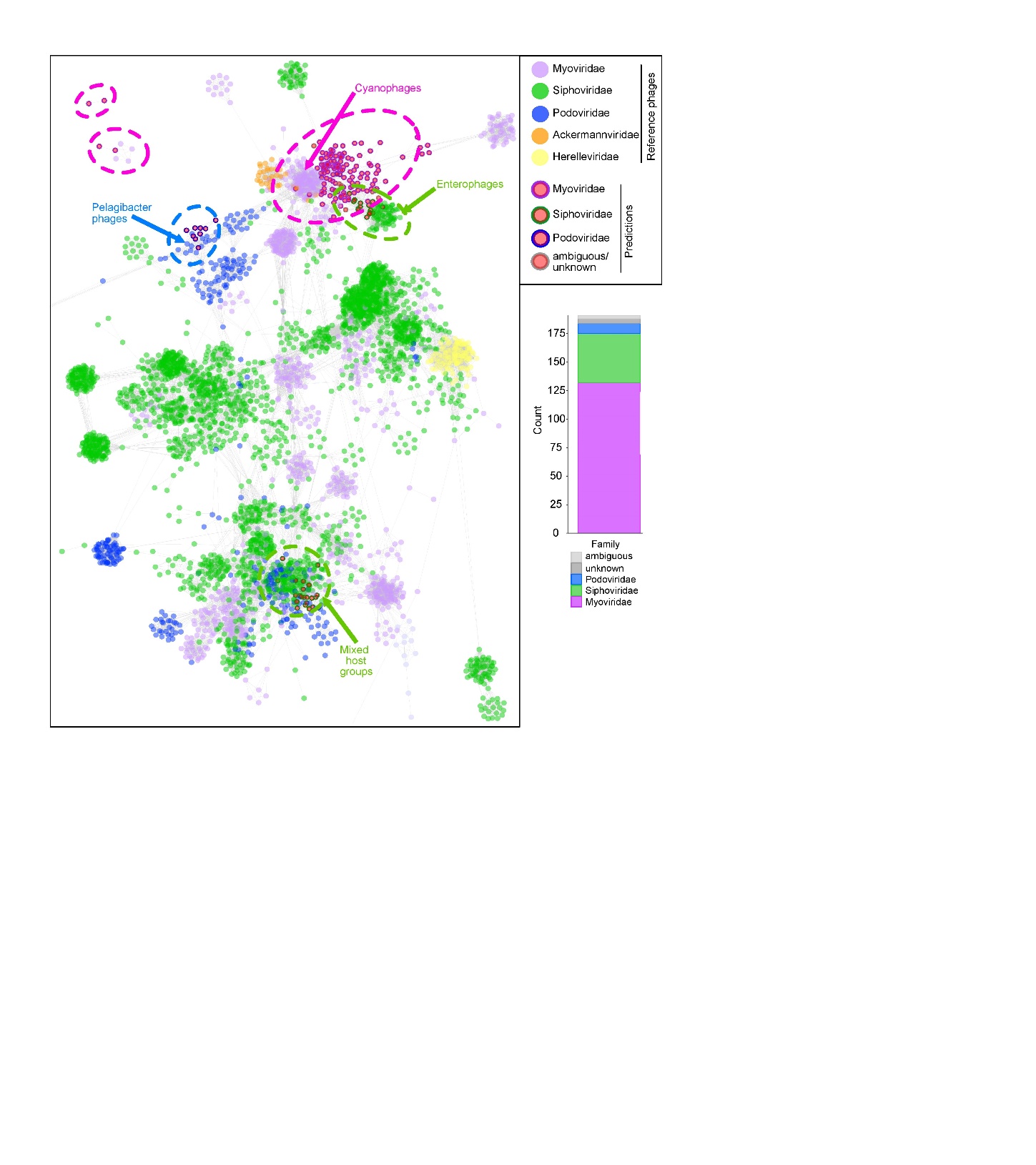
**硫噬菌体在*Caudovirales*目下分类呈现多样化**

**Sulfur phages are taxonomically diverse within the order *Caudovirales***

我们应用两种方法对已识别的mVC进行分类和集群。首先，我们使用参考数据库相似度搜索来将每个mVC分配给25个不同的原核生物感染病毒家族中的一个。大多数mVC为*Myoviridae*科（132个mVC；69%）、*Siphoviridae*科（43个mVC；22%）和*Podoviridae*科（9个mVC；5%）。这三个科代表属于*Caudovirales*目的dsDNA噬菌体。其余7个mVC被认定为不确定的冠状病毒（3 mVC；1.5%）和在目和科水平上均属未知（4 mVC；2%）的病毒。然而，根据这里提供的数据和以前的分类，7个未分类的mVC可能属于三个主要的Caudovirales病毒科之一（图4）。

根据这些结果，我们从NCBI基因库数据库中建立了一个具有参考病毒的病毒蛋白共享网络（图4）。mVC分为四个含有*Myoviridae*, *Siphoviridae* 和*Podoviridae*参考病毒蛋白的主要簇和4个分布在主要簇外单独的mVC簇。在7个预测模糊/未知的mVC中，有6个和*Myoviridae*和*Siphoviridae* mVC以及参考噬菌体聚集在一起，进一步表明它们与*Caudovirales*的主要科有关。总的来说，网络图验证了基于参考病毒的分类分配结果（即预测为*podoviruses*的mVC和参考的*podoviruses*聚集在一起，*myoviruses* 和*siphoviruses*也同样如此）。基于这些发现，我们假设DSM AMG在感染期间的功能很可能受到特定宿主硫代谢而不是病毒分类的限制。DSM AMG在*Caudovirales*中的广泛分布进一步表明，这种调节机制是在噬菌体的多个分类分支中建立起来的，可能是通过独立产生或者通过基因转移而获得。大多数mVC聚集在不同分类和宿主范围的参考噬菌体基因组中，尽管mVC和这些参考噬菌体之间没有足够显著的蛋白质相似性以表明它们在属水平上的相似性。鉴于*Pelagibacter*簇中包含了一个SUP05感染的mVC，宿主的范围很可能超出了这些被标示的类群。在参考数据库目前状态下，这种类型的蛋白质共享网络不能用于可靠地预测这些未培养的mVC的宿主范围，而是表明共享分类（例如，mVC *podoviruses*与参考*podoviruses*聚类在一起，*myoviruses* 和*siphoviruses*也同样如此）。基于系统发育和AMG蛋白的相似性，mVC的宿主范围似乎主要是来自SUP05/*Thioglobus*分支的γ-变形杆菌，同时有可能将宿主范围扩大到β-变形杆菌的*Methylophilaceae*科（图3）。通过使用对来自25个宏基因组的7178个spacers的CRISPR分析，我们无法验证假定宿主和任何mVC存在相互关系。

**图4** mVC的分类分配和具有参考噬菌体的蛋白质网络聚类。在蛋白质网络中，每个点代表一个单个的mVC（带轮廓的圆）或参考噬菌体（没有轮廓的圆），每个点通过线连接代表共享的蛋白质内容。具有更多相似性的基因组（即点）通过更接近和更多的联系来可视化。用虚线描绘的集群的注释是人工标注的。mVC分类根据自定义参考数据库和脚本的预测，被标上不同的颜色，在条形插入小图中显示。



**硫噬菌体在不同的环境表现出多样化和遗传嵌合体特性**

**Sulfur phages display diversification across environments and genetic mosaicism**

为了进一步评估已识别的mVC的多样性及其进化史，我们分析了单个mVC之间的共享蛋白组以及基因安排。所有来自94个mVC的预测蛋白质，不包括只编码类*tusE* AMG的mVC，都被聚集成蛋白质簇。尽管代表了部分基因组序列，我们研究mVC多样性的蛋白质聚类方法使用来自NCBI RefSeq的*Caudovirales*噬菌体进行了评估和验证。总共产生了794个代表3677个蛋白质的蛋白质簇，大致对应于单个蛋白质家族。只有少数蛋白簇在mVC中广泛共存，包括常见噬菌体蛋白（如PhoH、NifU、IscA、核酸酶、解旋酶、赖蛋白酶、RNA/DNA聚合酶亚基、ssDNA结合蛋白和形态特异性结构蛋白）（图5a）。由于部分mVC序列上的基因缺失，mVC之间缺乏共享的蛋白质簇这一现象是可以被预期的。然而，无论基因组完整性如何，不同的噬菌体谱系都很少共享蛋白质簇。总的来说，蛋白质分组的结果与分类分组的结果一致，进一步突出了编码DSM AMG的噬菌体基因组的多样性。缺乏普遍共享的蛋白质簇也表明DSM AMG的功能独立于其他宿主代谢途径，并可能严格补充宿主DSM途径。

为了确定共享的蛋白簇是否与DSM AMG相关，并进一步强调mVC的多样性，我们生成了第二组蛋白簇，对应于DSM AMG之前和之后的5个蛋白，同时也包括AMG。由于整个mVC真实的完整性不能确定，这个与DSM AMG相邻的11个蛋白的子集被用来可以认为是最好地代表潜在的共享特征，无论病毒基因的完整度情况如何。第二组包括70个mVC（我们排除了24个病毒，其中编码的DSM AMG在片度末端的5个基因内）。我们总共生成了116个蛋白质簇。有趣的是，我们得到与上面的分析结果几乎相同的广泛分布蛋白质簇，即PhoH、NifU、IscA、GrxD、TusA、NrdAB、RNA/DNA聚合酶亚单位、ssDNA结合蛋白和形态特异性结构蛋白。但是，共享蛋白质簇只代表所有簇中的一小部分子集。因此，除了铁硫簇形成之类的常见功能外，mVC还编码不同的蛋白质，这可能是由不同的进化背景引起的。

大多数根据整个mVC共享蛋白质簇形成分支的mVC可以通过共享的分类和/或环境来源来解释。这一观察结果进一步证实，尽管mVC只代表了部分序列，但它们编码了足够的信息，以便准确地分组为各个分支。也就是说，通过基因组排列得到的相似的mVC，与分类和环境来源一样，被发现分为同一个分支。例如，16个编码来自海洋环境编码*soxYZ*的*Myoviridae* mVC聚集在一起，它们之间只是各自具有代表性的蛋白质簇的总数有所不同（图5a）。然而也有例外，比如7个*dsrC*编码的mVC显示出多变的成对蛋白质相似性（在50%相似度截止水平上），其*dsrC*基因在其基因组中位置有所变化，尽管它们的其他基因有明显的共享和独特的基因共线性（图5b）。这7个mVC起源于三种不同的海洋环境类型（沿海、海洋和潮间带），它们都被预测为*myoviruses*（图5b）。这种多样性可能被解释为*dsrC*基因随着时间推移而保留，尽管基因组的组成部分正在经历基因交换、重组事件或突变积累。众所周知，噬菌体表现出基因嵌合特性，或基因和基因区域的交换和多样化。从不同的海洋环境（潮间带、咸水和浅海）中编码*soxYZ*的*myoviruses*（图5c）以及从热液环境中编码*dsrC*和*dsrA*的*siphoviruses*也可以得出相同的结论（图5d）。除了在不同的环境类别中分布，这些每个蛋白质共享分支基因嵌合的mVC在地理上也是广泛分布的（图5e）。此外，一个编码*soxYZ*的mVC（Ga0066606\_10000719）也编码同化硫代谢AMG *cysC*（图5b）。这提出了一个有趣的不常见现象，表明这种特殊的mVC，以及其他三个编码*cysC*（Ga0052187\_10001、Ga0052187\_10007和JGI24004J15324\_10000009）的mVC，同时针对异化和同化的硫代谢、更普遍地影响宿主的硫代谢。

Diagram

Description automatically generated

**图5** mVC蛋白质聚类和基因组比对图。（a）mVC分层蛋白质聚类。其中每一行代表一个蛋白质簇（共887个），每列代表单个mVC（共94个）。图中显示了编码的AMG、估计的分类、环境来源和每个mVC的蛋白质簇数量。（b）、（c）和（d）各自的分支由彩色虚线表示。来自不同环境的（b）7个不同的编码*dsrC*的*Myoviridae* mVC，（c）4个来自不同环境编码*soxYZ*的*Myoviridae* mVC，（d）来自热液环境的编码*dsrA*和*dsrC*的4个*Siphoviridae* mVC的基因组排列。对于基因组排列，每条黑线代表一个基因组，箭头代表预测的蛋白质，这些蛋白质根据VIBRANT的注释染色；基因组由代表tBLASTx相似度的线连接。（e）在b、c和d中描述的15个mVCs的地理分布图，图中标注了各自的分支、环境来源和分类科。

**基于组学数据分析估计硫噬菌体对硫氧化的贡献**

**Estimations of sulfur phage contributions to sulfur oxidation based on omics-data analyses**

我们利用包含mVC的宏基因组数据集来计算每个AMG噬菌体：总基因的覆盖度比率。在群落内的噬菌体：总基因的比率和每个预测的噬菌体-宿主对的噬菌体：总基因比率可以用来估计噬菌体对硫和硫代硫酸盐氧化/歧化代谢的贡献。这依赖于这样一种假设，即基因比可以成比例地反映真正的代谢活动，而且处于病毒感染细胞状态的宿主细胞与未受感染的微生物相比，保持着相同的环境适应度水平。通过将宏基因组reads覆盖到宏基因组内的AMG和假定的细菌宿主上，我们获得了mVC AMG与总基因的比率，这代表了AMG功能对具有代表性代谢，比如硫氧化的相对贡献。我们计算了mVC *dsrA*（图6a）和*soxYZ*（图6b）在热液湖、淡水湖和*Tara* Ocean海洋宏基因组数据集中的基因覆盖率。我们从DsrA和SoxYZ的系统发育树中鉴定了含有mVC AMG的噬菌体-宿主基因基因对。我们的结果表明，噬菌体*dsrA*在热液环境中的贡献主要来自SUP05γ-变形菌分支2；噬菌体*soxYZ*是环境位特异性的，在Croche湖、Fryxell湖和Tara Ocean海洋样本主要以β-变形菌分支、类*Methylophilales*分支和γ-变形菌分支为主。这表明AMG在每个环境中分布并在每个环境中发挥潜在作用的特异性。平均噬菌体：总基因覆盖率也不同，噬菌体*soxYZ*：总基因在*Tara* Ocean海洋样本的比率最高（34%），其次是噬菌体*dsrA*：总基因在热液样本的比率（7%），最后是噬菌体soxYZ：总基因在淡水湖的比率（3%）。在海洋中的噬菌体*soxYZ*（硫载体基因）比在另外两种环境中的*dsrA*（Dsr复合物的催化核心组成部分编码基因）具有更高的噬菌体：总基因比率。这里使用的*Tara* Ocean海洋样本都来自表层远洋地区，属于低浓度硫的含氧层，而羽流样本（Lau Basin）来自含高浓度硫的深海热液生态系统。然而，根据与噬菌体*dsrC*相关的观察，我们的结果表明，编码硫载体而不是催化亚基的AMG似乎更受噬菌体的青睐。这些发现是出乎意料的，因为我们期望远洋环境比热液羽流环境具有更低的硫氧化活性。虽然这里研究的有限的环境类型、条件和硫AMG并没有提供足够的统计可信度来推广这些结果，特别是在比较来自不同环境的不同基因时，但是噬菌体中高丰度的硫载体基因仍然是一个常见的现象。此外，虽然基因丰度比并不一定代表功能的贡献，但这种情况仍然提供了一个合理的估计，以表明噬菌体硫AMG有相当大的硫氧化贡献作用。

随后，我们计算了个体噬菌体-宿主对的噬菌体：宿主AMG覆盖度比率，以估计噬菌体AMG在每个环境样本中的潜在功能贡献（图7a、b）。平均计算SUP05分支1和SUP05分支2中*dsrA*噬菌体-宿主对覆盖度比率以及淡水湖和*Tara* Ocean海洋样本中*soxYZ*噬菌体-宿主对覆盖度比率发现每对噬菌体：总基因覆盖度比率一般高于~50%。这些对内噬菌体：总基因比率远高于上述噬菌体：总基因在整个群落水平的比率。*Tara* Ocean海洋样本在这三种环境中也具有最高的噬菌体-宿主对内平均噬菌体：总基因比率，这与上面提到的整个群落的比率模式一样。为了估计群落中病毒感染细胞的百分比，我们使用平均值16%和15%来代表海洋（其中自由生活和颗粒相关生活细菌的百分比范围分别为3-31%和3-26%）和淡水湖泊（百分比范围为1 到 17%+/-12%)）中的病毒感染细胞占总细胞比例。估计的整个群落内噬菌体：总基因的覆盖度比率应该为病毒感染细胞百分比乘以噬菌体-宿主对内的平均噬菌体：总基因覆盖度比率（作为噬菌体基因覆盖度），然后除以总基因覆盖度。我们发现估计的比率与观测到的比率不一致（*Tara* Ocean：估计的比率68% vs观测到的比率34%；热液环境：估计比率15-38% vs观测比率 12-20%；淡水湖：估计比率27% vs观测比率3.1%）。这可能是由于部分宿主细胞受到不含DSM AMG的噬菌体感染，因为这些病毒感染细胞不贡献噬菌体硫代谢基因；或者也可能是由于病毒感染细胞状态的细胞百分比低于平均水平。

上述分析表明，DSM AMG可能在群落水平和单个噬菌体-宿主对的水平上对宿主驱动代谢的功能有着显著贡献，而每个环境和每个生态位特异性的AMG的贡献比率有很大差异。更重要的是，噬菌体编码的*soxYZ*对远洋海洋微生物群落有很高的基因覆盖度贡献，这突出了噬菌体驱动的硫循环代谢和整个硫代硫酸盐氧化/歧化代谢的生态功能意义，也显示这一方面的研究目前仍然还是很缺乏。

A picture containing diagram

Description automatically generated

**图6** 噬菌体与总*dsrA*和*soxYZ*基因覆盖度比率。（a）噬菌体*dsrA*与总（噬菌体和细菌*dsrA*基因一起）基因覆盖度比率。来自不同γ-变形菌SUP05分支的噬菌体*dsrA*基因的贡献用不同的颜色标注。平均噬菌体*dsrA*：总基因比率的计算来自于12个样本。（b）噬菌体*soxYZ*与总基因覆盖度比率。来自三个不同分支的噬菌体*soxYZ*基因的贡献用不同的颜色标注。淡水湖和*Tara* Ocean海洋样本的基因分别进行比较，淡水湖和*Tara* Ocean海洋样本的平均噬菌体*soxYZ*：总基因覆盖度比率分别进行比较。*Tara* Ocean海洋样本的ID被标记在图中，相应的宏基因组的ID被列在补充数据4中。

**硫噬菌体*dsrA*活性随着地化学梯度快速变化**

**Rapid alteration of sulfur phage *dsrA* activity across geochemical gradients**

由于DSM AMG与微生物关键的能量产生代谢有关，我们想研究硫噬菌体对变化的地化环境条件的反应，其中涉及到病毒感染细胞驱动的生物地球化学循环能力。在热液生态系统中，还原性化学底物如硫化氢、硫、甲烷和氢从高温热液喷口释放出来，与冷海水混合后迅速稀释，形成急剧变化的化学梯度。深海环境中的微生物通过提高其在热液环境中的代谢活性，对高浓度还原性硫化合物作出反应。这些特征使热液和背景深海环境成为研究AMG表达变化的对比生态位。我们使用转录组谱来研究Guaymas Basin的热液喷口和加州湾的背景深海样本中的噬菌体：宿主基因对的表达情况。硫噬菌体的表达（单位RPKM）在深海中为0.03-3而在热液环境中为0.40-39。平均噬菌体*dsrA*在热液体与深海背景的表达比率为15。受基因库及其生物学特性的限制，噬菌体本身没有能力独立感知和对硫化合物做出反应。然而，我们的结果表明，发生在病毒感染细胞内的硫噬菌体活性与地球化学变化密切相关，还原性硫化合物浓度较高的环境中观察到其具有较高的活性。

在Guaymas Basin热液环境中，如两对SUP05分支1噬菌体和宿主*dsrA*基因所反映的，噬菌体与宿主*dsrA*转录比范围为0到0.11（图7c）。相比之下，在Chesapeake湾，正如两对噬菌体和宿主*dsrA*转录本（Chesapeake湾*dsrA*分支）所反映的那样，噬菌体与宿主*dsrA*转录本比率从1.9到无穷大不等（宿主转录本丰度为零）。热液宏转录组中噬菌体*dsrA*的低丰度与热液宏基因组中噬菌体*dsrA*的高丰度（在Guaymas Basin和Lau Basin中）形成鲜明对比（图7a，c）。对这一观察结果的一种解释是，这种情况可能是一种意外，但并不代表热液系统中真实的噬菌体基因表达模式，可能发生在取样前噬菌体活性非常高的情况下。在这种情况下，大多数宿主/病毒感染细胞可能在病毒感染后被裂解。

Diagram

Description automatically generated

**图7** 噬菌体与宿主*dsrA*和*soxYZ*基因的覆盖度比率以及噬菌体与宿主对之间*dsrA*基因表达的比较。（a）每个噬菌体-宿主对的噬菌体*dsrA*：总基因覆盖度比率。在SUP05分支 1和2中噬菌体宿主对的平均噬菌体*dsrA*：总基因比率分别由5和11对基因对来计算。（b）每个噬菌体-宿主对的噬菌体*soxYZ*对总基因覆盖度比率。来自三个不同分支的噬菌体*soxYZ*基因的贡献由不同的颜色标记。淡水湖和*Tara* Ocean中的噬菌体-宿主对的平均噬菌体*dsrA*：总基因比率分别计算。*Tara* Ocean海洋样本的ID被标记在图中，相应的宏基因组的ID被列在补充数据4中。（c）在Guaymas Basin宏转录组中噬菌体和宿主*dsrA*基因的表达比较。热液和背景海洋宏转录组数据集覆盖在同样的数据库上。（d）在Chesapeake湾宏转录组中噬菌体和宿主*dsrA*基因的表达比较。全部Chesapeake湾宏转录组数据集覆盖在同样的数据库上。基因表达水平是以基因序列丰度和基因长度标准化后的RPKM作为单位的。

**讨论**

自从首次描述使用AMG的病毒代谢重新编程以来，人们一直对病毒辅助代谢对全球能量流动和生态系统中营养可用性的程度和总体影响感兴趣。通过对宏基因组的调查和研究，我们扩大了目前对影响异化性硫氧化过程病毒辅助代谢的理解。具体地说，我们已经证明了不同的噬菌体谱系参与了这些过程，我们研究了它们的生物地理特征、生态特征和进化历史，并评估了它们对微生物组的潜在影响。由此，关于病毒辅助代谢和硫循环的几个假设和问题得到解释。

首先，我们的发现支持了先前的假设，即病毒代谢针对宿主代谢途径的关键或瓶颈步骤。DsrA、DsrC、SoxYZ、SoxC和SoxD都缓解了硫和硫代硫酸盐氧化/歧化代谢的瓶颈。我们没有识别硫氧化途径中的其他基因，如硫化物：醌氧化还原酶、黄细胞色素*c*细胞色素/黄素蛋白亚基、APS还原酶亚基、硫酸盐腺苷酸转移酶、*dsrB*或*soxAB*，以进行硫氧化的其他必要步骤。然而，这也提出了另一个问题，即为什么DsrB，与Dsr A构成二聚体的亚基，还没有被发现可以是AMG。此外，硫载体蛋白，而不是酶，似乎更受噬菌体的青睐。本研究共有174个mVC编码至少一个硫载体（*dsrC*、类*tusE*、*soxYZ*）蛋白，其余17个编码酶催化亚基（*dsrA*、*soxC*、*soxD*）。在噬菌体群落中，噬菌体硫载体比*dsrA*等催化亚基更丰富。这可能是由于更需要硫载体（例如*dsrC*）来驱动异化性硫转化。这一假设的证据是，与各自的催化亚基（如*dsrA*）相比，硫载体通常在宿主细胞中有组成型的表达。通过在感染期间内提供这些重要途径成分的转录本和蛋白质，编码DSM AMG的噬菌体可能会从病毒感染细胞内获得更大的能量并且从自催化底物中获益更多。

通过mVC蛋白聚类和基因组对齐所显示的数据（图5）支持了DSM AMG能在快速进化的噬菌体基因组上保留的假设，这特别指出了AMG在增加噬菌体复制能力和适应度方面的作用。虽然大多数mVC中AMG的分散机制尚不清楚，但很可能基于在各自基因组中AMG位置的相似基因安排，在每个分支内发生了单一的AMG转移事件。这表明，尽管单个mVC种群存在生态位（即地理和环境）差异，AMG仍然会保留。据推测，AMG和其他噬菌体基因一样，必须提供显著的适应度优势，才能随着时间的推移保留在不断进化的噬菌体基因组上。

综上所述，这些观察结果支持了这样的结论，即病毒辅助代谢针对宿主代谢途径的关键步骤，以进行微调的、宿主依赖性的能量产生或营养获取的操作。虽然DSM AMG的适应度效应尚未在模型系统中被量化，但尽管编码能力受到限制，但已识别的mVC的地理分布和噬菌体AMG的保留强烈地表明，编码DSM AMG具有显著的适应度优势。如果没有噬菌体-宿主对的培养物和随后的遗传操作能力，对于编码DSM AMG所获得的量化的适应度优势研究仍然难以实现。此外，一个模型系统将有助于阐明AMG的功能，这将会超出了这里提出的蛋白质结构域分析的证据。虽然大多数AMG编码保守的功能域和残基，但不同序列的识别，如类*tusE* AMG编码一个单一的半胱氨酸残基或*soxD* AMG似乎缺乏一个细胞色素*c*基序，需要对AMG编码的蛋白进行进一步的生化实验评估。例如，发现蓝细菌噬菌体编码的变异PebA排除了后续宿主酶PebB的必要性，短路了原宿主途径。硫代谢的AMG编码的变异蛋白质也可能使得宿主途径短路，或者使用变异AMG增加宿主硫氧化途径的速率。

由于DSM AMG已经在所有三个主要*Caudovirales*所属科上被识别出来，所以AMG带来的适应度优势可能和特定处理硫氧化和提升电子产量途径的速度或效率相关，而不是与噬菌体分类相关。基于编码光合作用AMG的蓝细菌噬菌体系统的证据表明，硫AMG在提高代谢途径的速度或效率方面的一个潜在效用是提高用于基因组复制dNTPs的生产收益率。进一步的证据表明，AMG也可以上调编码不稳定蛋白产物的重要代谢基因的表达。在这种情况下，宿主细胞适应于这种代谢限制，但噬菌体的复制速率直接取决于给定的AMG蛋白产物的翻译速率。因此，噬菌体而不是宿主，受益于代谢基因的额外拷贝，从而导致噬菌体基因组上AMG的重组和保留。由于裂解性噬菌体基因组上高丰度的AMG，噬菌体主要在短期的活性裂解细菌感染期间获益。然而，同化硫酸盐还原基因（即*cysC*）与DSM基因的共存提供了一个更普遍的硫操作的可能的例外的例子，强调了进一步研究病毒辅助代谢的必要性。

基于噬菌体：总基因覆盖度比率测量的宏基因组和宏转录组中噬菌体DSM的丰度表明，噬菌体介导的还原性硫转化可以对群落内硫的通量和预算有非常重要的贡献（图8）。在每个噬菌体-宿主对中，噬菌体基因贡献了与硫和硫代硫酸盐氧化途径相关的基因覆盖度的一半以上，这突出了编码DSM AMG的噬菌体在重塑硫循环中的作用，特别是对还原性硫的氧化作用。还原性硫化合物，如硫化氢、硫、硫代硫酸盐在Guaymas Basin的热液系统中含量很高，其中硫化氢浓度高达~6 mmol/kg（热液口端测量），而在背景海水中其浓度则可以忽略不计。先前报道的对Guaymas Basin热液系统中硫氧化细菌的能源预算的估计表明，微生物代谢产生的能量可达~3900 J/kg，其中高达83%可能来自硫氧化。与深海背景相比，硫噬菌体*dsrA*表达水平（来自病毒感染细胞）在热液系统中升高，这暗示着病毒感染细胞介导的噬菌体驱动的硫氧化对整体能源预算有重要贡献。假设在Guaymas Basin热液环境中，噬菌体：总*dsrA*基因覆盖度比率为10%（这是Lau Basin热液环境的平均水平），可以估计热液流体中微生物代谢的~320J/kg能量实际上可能是通过硫AMG转化的。虽然大多数噬菌体的宿主代谢操纵和裂解可能发生在噬菌体没有携带AMG的情况下，但我们的研究表明，编码硫AMG的噬菌体可以是硫生物地化循环的直接组成部分，能够操纵与多种还原性硫化合物相关的微生物代谢。这种直接的代谢操纵可能会影响生态系统规模上的硫代谢预算。因此，未来对生物地化循环的评估必须纳入噬菌体的作用及其对硫汇的影响。受本研究中基于组学的方法的限制，我们无法实现更精细水平的噬菌体-宿主相互作用和活性的研究，这证明了未来在宿主细胞内加强精细水平的噬菌体AMG活性研究是非常必要的。

在地球上各种不同的环境中，还原性硫汇包括深海或地下沉积的硫化铁来源以及异化性硫酸盐还原和有机硫矿化产生的还原性硫（图8a）。硫噬菌体AMG辅助代谢有助于硫产生的能量重新分配，并可以改变其预算。到目前为止硫能量循环只归因于微生物过程而还尚未涉及到噬菌体（图8a）。在病毒感染细胞中，噬菌体介导的硫氧化将利用硫代谢途径的基因表达转录本，并产生酶为噬菌体复制重定向能量（图8a）。全球分布的硫噬菌体广泛分布在各种环境中，对硫汇以及营养物质和能量循环产生了重大影响（图8a）。同时，噬菌体AMG介导的硫氧化可以短路化微生物硫循环从还原硫汇到溶解和颗粒有机物（DOM/POM）（图8b）。如果没有病毒感染，还原性的硫汇产生的能量通常会被用于初级生产，以促进微生物细胞的生长，然后在食物链上转移到采食性动物身上。通过细胞排泄效应、细胞死亡和营养物质的释放，由硫的初级生产产生的DOM/POM将被释放到环境中。然而，在硫噬菌体感染期间，还原性硫汇在病毒感染细胞中产生的能量可以用于噬菌体的再生和繁殖。在病毒粒子的产生和包装后，裂解性噬菌体会裂解宿主细胞，并将DOM释放到环境中。这种DSM AMG介导的路径从而短路化微生物硫循环。此外，由还原性硫氧化过程产生的POM也可以被隔离在地下深处沉积的碳汇中。目前尚不清楚噬菌体将如何以及在多大程度上改变隔离碳和生物可用碳之间的碳循环格局，而可以肯定的是，在未来的全球生物地化循环和气候变化中噬菌体AMG代谢引起的变化应该得到明晰的阐述。

总之，我们描述了与硫相关的噬菌体辅助代谢的分布、多样性和生态学特征，并证明了硫噬菌体在环境中的丰度和活性。然而，许多问题仍没有得到回答。未来的研究将包括阐明硫噬菌体和宿主相互作用的机制，在单个病毒感染细胞、微生物群落和生态系统尺度上的硫代谢的重塑，以及受硫噬菌体影响的硫预算的限制和分配。

Diagram

Description automatically generated

**图8** AMG在硫代谢中的生态和功能概念图。（a）DSM AMG对还原性硫转化预算的影响。（b）营养循环中噬菌体介导的代谢短路化微生物硫循环示意图。

**参考文献**

Kieft, K., Zhou, Z., Anderson, R.E. et al. Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. Nat Commun 12, 3503 (2021). https://doi.org/10.1038/s41467-021-23698-5

**作者简介**

**Zhichao Zhou（周之超）**

Zhichao obtained his Ph.D. from The University of Hong Kong in 2017, then he moved to The Chinese University of Hong Kong and University of Wisconsin-Madison, US, studying sympatric speciation of coastal mangrove rhizospheric microbial populations, microbial genome diversification and adaption, and microbial genome-based bioinformatic tool development. He is skilled in metagenome and metatranscriptome analyses and shows broad interests in human and environmental-related microbiome and virome that are significant to public health and environmental protection and maintenance. Zhichao’s research interests are mainly focused on metagenome and metatranscriptome based biogeochemical functions and activities, environmental and energy dynamics mediated by microorganisms and viruses, microbial community interactivities, domestic environments influenced by anthropogenic activities, resistome and virulome, populational genomic-based microbial community characterizations and microbial evolution mechanisms.

**Karthik Anantharaman**

Karthik Anantharaman is an Assistant Professor in the Department of Bacteriology at the University of Wisconsin-Madison. He received his Ph.D. in Earth and Environmental Sciences from the University of Michigan in 2014 studying the geomicrobiology of deep-sea hydrothermal plumes. His interdisciplinary research program focuses on understanding the cycling of sulfur and nutrients, with a strong emphasis on the microbial and viral processes that transform them in environmental and human systems. Karthik’s research uses a combination of cultivation-independent, cultivation-dependent, and field-based approaches to investigate the interplay between the microbiome, virome, and chemistry in hydrothermal plumes, oxygen minimum zones, and the deep oceans. He just received NSF CAREER award this year for the research of developing new approaches to study viral ecology and conduct virus-centric bioinformatics courses and workshops.

Home page: https://anantharamanlab.wisc.edu/