Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Antimicrobial Resistance—Article

ARGs-OAP v3.0——抗生素抗性基因数据库的更新和分析流程升级

殷晓乐，郑夏婉，李丽观，章安妮，姜小涛，张彤*

Environmental Microbiome Engineering and Biotechnology Laboratory, Center for Environmental Engineering Research, Department of Civil Engineering, the University of Hong Kong, Hong Kong 999077, China

ARTICLE INFO

Article history:
Received 4 May 2022
Revised 31 August 2022
Accepted 12 October 2022
Available online 27 December 2022

关键词
SARG 数据库
ARGs-OAP
抗生素抗性基因
环境宏基因组
量化

摘要

由抗生素抗性基因（ARG）编码的抗生素抗性激增，对全球公共卫生构成日益严重的威胁。随着技术的进步，特别是在宏基因组测序的普及方面，科学家们已经获得了高速解读不同样本中ARG谱的能力。为了以高通量的方式分析数千个ARG，需要标准化和集成的流程。广泛使用的抗生素抗性基因在线分析流程（ARGs-OAP）的新版本（v3.0）对参考数据库——结构化抗生素抗性基因（SARG）数据库和综合分析流程都进行了重大改进。SARG通过序列管理得到加强，从而提高注释的可靠性，纳入新出现的抗性基因型，并确定严格的机制分类。该数据库以树状结构的形式在线组织和可视化。针对不同的应用程序场景将它划分为不同的子数据库。此外，ARGs-OAP已经通过调整量化方法、简化工具实施和用户自定义参考数据库的多种功能进行了改进。而且，该在线平台现在提供了一个多样化的生物统计分析工作流程和可视化软件包，用于有效解读ARG图谱。ARGs-OAP v3.0具有改进的数据库和分析流程，将有利于学术界、政府管理部门和有关ARG环境流行率风险评估工作。

1. 引言

抗生素的发现改变了临床治疗领域，并挽救了受到传染病威胁的数亿人的生命。然而，抗生素的滥用和误用随后导致了后抗生素时代全球对抗生素耐药性（AMR）的关注。许多国家已经共同努力推进抗生素管理，并加强对抗生素抗性基因（ARG）[1–2]的监测。基因组测序在ARG监测中的应用越来越广泛，在高通量分析和解码ARG基因组背景方面具有优势。随着测序技术的发展和威胁人类健康的AMR危机，越来越需要可靠的参考数据库和生物信息学工具，以便利用DNA序列的大数据对ARG进行快速、准确的注释、分类和定量[3]。

结构化ARG（SARG）数据库于2016年首次发布，迅速成为广受欢迎的ARG数据库之一。它是基于综合抗生素耐药数据库（CARD）[4]和抗药基因数据库（ARDB）[5]构建的，生成包含4049个变异的集合，该集合具有类型-亚型序列层次结构，并且每个变异都有明确的分类序列[6]。ARG类型代表抗性基因编码的蛋白质所对抗的抗生素（如一些研究中使用的抗生素/药物类别），而亚型代表基因的基因型（如一些研究中使用的ARG家族）。2018年，对SARG数据库进行了进一步扩展，将其发展为v2.0版本，其中通过序列比对和关键字匹配等严格选择

* Corresponding author.
E-mail address: zhangt@hku.hk (T. Zhang).

英文原文：Engineering, 2023, 27(8): 234–241
引用本文：Xiaole Yin, Xiawan Zheng, Liguan Li, An-Ni Zhang, Xiao-Tao Jiang, Tong Zhang. ARGs-OAP v3.0: Antibiotic-Resistance Gene Database Curation and Analysis Pipeline Optimization. *Engineering*, https://doi.org/10.1016/j.eng.2022.10.011

标准，从美国国家生物技术信息中心（NCBI）非冗余（NR）数据库中纳入了更多经过匹配的ARG参考序列。抗性基因在线分析流程（ARGs-OAP）可用于ARG注释、分类和两步分析定量。第一步是通过Usearch对ARG序列进行快速过滤[7]，第二步是使用基本的局部比对搜索工具（BLAST）进行精确分类[8]。ARGs-OAP得到了全球关注，越来越多的用户推动了该工具的不断改进，随着SARGfam的部署和基本单拷贝标记基因在细胞数量定量中的应用而被更新到v2.0版本[9]。

需要不断改进基于SARG的ARGs-OAP，以提高其性能及与其他下游分析的集成。因此，本研究描述了ARGs-OAP v3.0的最新更新，如图1所示。其中包括：①一个精心策划的数据库，通过修订层次结构减少注释偏差；②升级了注释、分类和量化工具，增加了环境样本的ARG覆盖率，以及计算ARG丰度的新方法；③改进网站，对ARG进行综合深入分析和统计可视化。

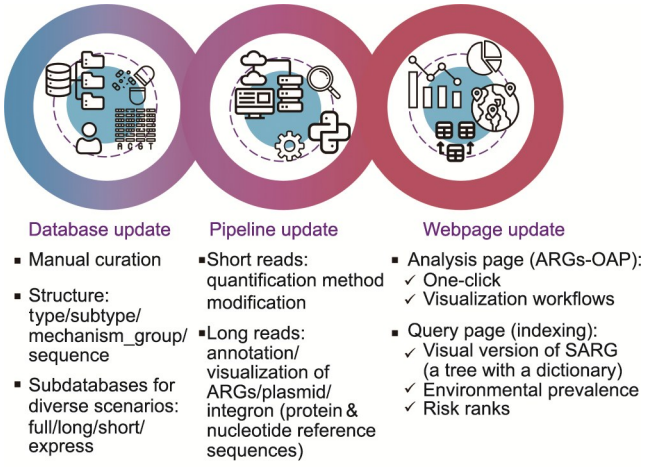


图1. ARGs-OAP v3.0已经更新，包括一个新的数据库、一个完善的流程和具有多种功能的网页。

2. 方法

2.1. 数据库管理

使用内部脚本对所有参考序列进行了严格的管理，然后通过参考文献†、分子专家、其他相关数据库和NCBI注释进行手动验证。对单个序列进行序列比对和关键字匹配，如果比对结果与关键字搜索匹配，则得到准确的分类。具体来说，首先，根据最新的知识，如四环素和大环内酯-林可酰胺-链霉素（MLS）抗性基因[10–11]的术语，筛选出特定抗生素类型的ARG。其次，ARG亚型的名称和分类由其他数据库进行补充，包括CARD数据库（v3.2.4，2022年7月27日下载）[12]。经过手动过滤，CARD数据库中的4641个序列中有713条被纳入SARG数据库，提供了更新的ARG亚型术语集合（表1）。此外，根据已发表论文中的分类，对SARG中的其余参考序列进行了单独审查。那些没有可用亚型/类型分类的序列被移除，以避免在定量中潜在的错误注释和误报。最后，将SARG数据库与NCBI NR数据库（2022年8月28日下载）进行比对，以检索更多的参考序列，并遵循严格的选择标准[9]。

表1 SARG v1.0/v2.2/v3.0和CARD v3.2.4数据库中的类型/机制/亚型/序列的计数

Item	Database			
	SARG v1.0	SARG v2.2	SARG v3.0-F	CARD v3.2.4
Type	24	24	32	–
Mechanism	–	–	11	–
Subtype	1 208	1 244	2 842	–
Sequence	4 499	12 085	13 672	4 641

SARG v3.0-F: SARG v3.0 full version.

2.2. 用于评估ARGs-OAP v3.0的模拟数据集

为了评估ARGs-OAP v3.0的性能，我们使用定制的脚本从Swiss-Prot数据库中生成了模拟数据集[13]。Swiss-Prot数据库（2020年4月20日下载）中关键词为“抗生素抗性”的序列被视为ARG，而Swiss-Prot数据库中的其他序列被视为非ARG。对整个集合生成了50个、67个和100个氨基酸（aa）蛋白序列的*k*-mers，以代表读长为150个、201个和300个碱基对（bp）的宏基因组数据集。当将ARGs-OAP应用于模拟数据集时，评估了不同截断点（即*E*值、相似度和比对长度比）。采用附录A中总结的计算方法，根据Matthews相关系数（MCC）、灵敏度和精度，评估了采用不同截断点的工具的鲁棒性。

2.3. 用于评估不同版本的ARGs-OAP的数据集

为了评估数据库更新带来的ARG丰度和多样性的变化，对来自7种不同环境类型的36个样品进行了量化分析，包括4个河水样本、3个沉积物样本、4个来自污水处理厂（WWTP）的厌氧消化污泥（ADS）样本、9个来自污水处理厂的活性污泥（AS）样品、两个来自污水处理厂的废水样本、两个来自污水处理厂的进水样本、12个来自牲畜粪便或养猪场的废水样本。对于每种环境类型，对从不同的宏基因组中量化的ARG类型的丰度进行平均，以代表该环境类型。

† https://smile.hku.hk/ARGs/Indexing.

2.4. 参考ARG的风险排序

参考ARG的风险排名框架基于Zhang等[14]的工作，该工作根据三个标准通过决策树将SARG数据库中的参考序列分为4个风险等级（等级I、II、III和IV）。第一，使用默认截断值在全球宏基因组集合（n = 1427，2022年9月17日之前获得的数据）和Refseq基因组集合（n = 256 788，2022年8月26日下载）中搜索SARG v3.0数据库中的所有参考ARG。那些未在任何宏基因组中检测到的参考序列被列为“未评估”。第二，将与人类相关的环境（包括人类粪便、牛粪便、猪粪便、污水、污水处理设施、农业领域、工业废水处理设施和矿山）中的ARG流行率与未受影响环境（包括海水、天然水、天然沉积物和天然土壤）中的丰度进行了比较（附录A中的表S1）。那些在人类相关环境中未被发现富集（定义为大于或等于100倍）的ARG被归类为“等级IV”。第三，在人类相关环境中富集的ARG中，通过搜索移动遗传元素（MGE）数据库（2022年4月4日下载的Refseq质粒数据库）判断的非移动ARG被归类为“等级III”。第四，在可移动的和与人类相关的ARG中，那些非由病原体携带的ARG被归类为“等级II”。最后，那些符合所有三个标准的参考ARG，包括①在人类相关环境中富集的、②可移动的和③病原体携带的，被归类为“等级I”，这表明风险最高。

2.5. 信息技术

ARGs-OAP v1.0 [6]和v2.0 [9]是基于Galaxy项目[15]部署的。在更新的版本中，Galaxy项目是通过定制的Python Flask框架、Vue.js框架和Quasar框架以及由R-Studio生成的支持性数据集进行原位开发的。MySQL存储支持使用大量数据的数据库索引，并由markmap包进行可视化。

3. 结果和讨论

3.1. SARG数据库更新

与SARG的前两个版本一样，SARG v3.0中的ARG参考序列以分层结构（类型-亚型-序列）组织，这有利于自上而下地解读环境样本，特别是当应用ARGs-OAP来量化ARG的表型（ARG类型）和基因型（ARG亚型）时。在SARG v3.0中，确定了抗性机制，形成了一个新的四层（类型-机制-亚型-序列）结构。其包括6个机制组：抗生素靶点改变、抗生素靶点保护、抗生素靶点替代、外排泵、酶失活和渗透性降低[16–26]。某些机制组被进一步划分为亚组。例如，外排泵进一步分为五个亚组：三磷酸腺苷（ATP）结合盒（ABC）转运蛋白、主要促进剂超家族（MFS）转运蛋白、多药和毒性化合物挤压（MATE）转运蛋白、抗性结瘤细胞分裂（RND）转运蛋白、小多药耐药性（SMR）转运蛋白[24–25]。

此外，在SARG v3.0中，对那些具有编码抗生素抗性的双组分系统或三组分系统的ARG亚型给予了特殊的“双组分”和“三组分”标签。例如，四环素抗性[27]需要一对具有外排泵[*tetA*(46)和*tetB*(46)]的基因。AcrEF-TolC是RND转运蛋白亚家族的三组分系统的另一个例子，其功能需要膜融合蛋白（AcrE）、内膜转运体（AcrF）和外膜因子（TolC）的参与[28]。

此外，对ARG类型和亚型的整理导致1717个新的ARG亚型被添加到SARG数据库，包括157个氨基糖苷、230个β-内酰胺、35个氯霉素、96个MLS、99个多药、106个喹诺酮、73个万古霉素和其他耐药亚型（表1和附录A中的表S2）。已经确定了11个同义词。特别是，SARG数据库现在除了CARD外，还包括127个ARG亚型，包括*mdtL*、*SHV-112*、*SHV-39*和*tetX1*。对于SARG v3.0数据库，已经对这些亚型名称进行了手动管理。

综上所述，由于对特定类型或亚型的分类不一致，从SARG v2.2中删除了1425个序列，并添加了3012个序列，从而更新了数据库SARG v3.0完整版（SARG v3.0-F），包含32个类型、2842个亚型和13 672个序列（表1和表S1）。

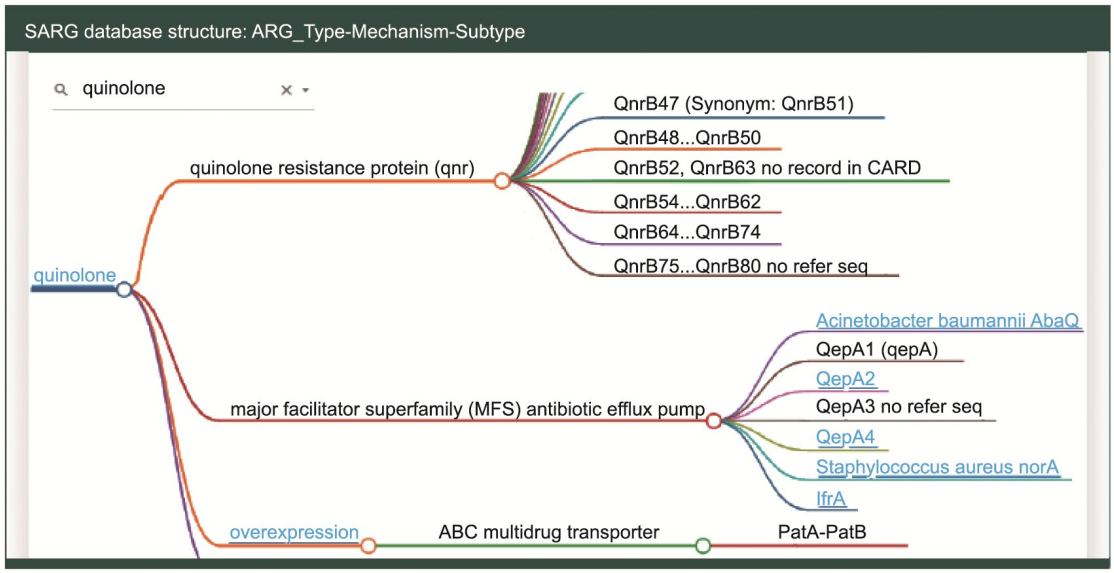
为了更好地利用不同长度的DNA序列对ARG进行注释和分类，在SARG v3.0-F基础上构建了SARG v3.0-L（长读数注释子数据库，n = 13 439），删除了233个由突变进化而来或在过表达时发挥作用的序列，这些序列不适合基于相似性搜索进行注释。此外，还创建了SARG v3.0-S（用于短读数定量，n = 12 746），作为SARG v3.0-L的子数据库，该数据库排除了用转录调控因子（包括激活因子和抑制因子）标记的693个序列，这些序列不能使用短读长进行正确注释。子数据库SARG 3.0-E（用于表达分析，目前为n = 10 538）只包括在不同环境的全面调查中至少检测到一次的环境流行SARG序列。SARG v3.0-S和SARG v3.0-E都是ARGs-OAP中的参考数据库，用于使用相似性搜索算法对环境宏基因组短读数数据集进行完整或快速分析。

3.2. 数据库索引平台

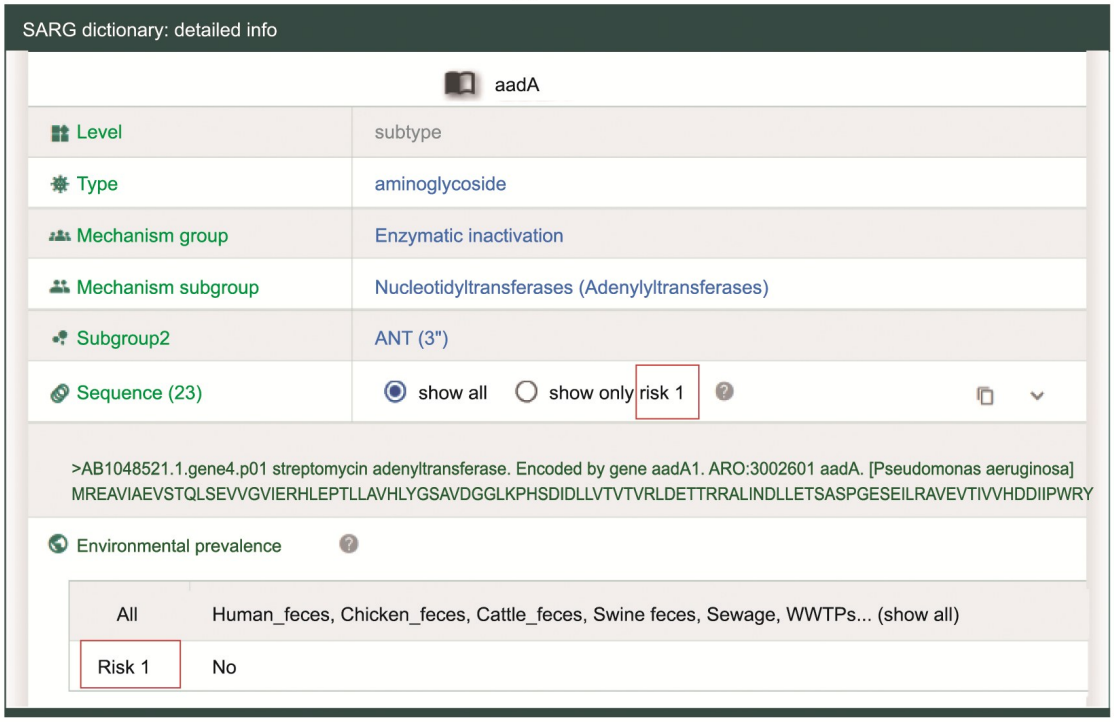
SARG v3.0-F数据库的结构清晰地显示在ARGs-OAP网站上，供用户检索每个基因和参考序列的信息，并在13 672个参考序列中引用任何感兴趣的ARG。

对SARG数据库的索引采用了两种格式（图2）。一种是分层树视图，其中每棵树植根于一种ARG类型，然后作为分支成长为不同的抗性机制、蛋白质家族和ARG亚型。树状结构索引是对SARG数据库的层次结构的一种用户友好的可视化。另一种显示格式被设计为将SARG数据库中的每个本体进行存档并提供全面的描述，类似于其他数据库，如UniProt [29]和Pfam [30]。

ARGs-OAP索引中最值得注意的部分是每个ARG亚型的环境流行率信息，这是由来自不同环境样本的1000多个宏基因组数据集的数据挖掘结果总结的（表S1）。此外，SARG v3.0-S中的12 746个参考序列被分为风险等级I、II、III和IV[†]。基于我们最近发表的风险等级方案[14]，风险等级I的ARG由于其跨越系统发育边界的高流动性、在人类活动中的广泛传播以及其宿主的致病



(a)



(b)

图2. SARG v3.0-F（n = 13 672）的可视化版本有两种格式。（a）具有支持搜索功能的树ARG类型；（b）SARG数据库中每个本体的存档信息，包括相关类型、亚型、机制组、亚组、参考序列和环境流行度信息。aadA：氨基糖苷-3'-腺苷酸转移酶。

[†] https://smile.hku.hk/ARGs/Indexing/riskranking.

性而最受关注。因此，风险等级I的ARG在网页上的字典中的“环境流行率”和“序列”部分都被突出显示。一般而言，ARG流行率数据和等级排名计划为学术界和政府了解ARG的传播和制定控制策略提供了有价值的参考。

3.3. ARG量化工具和可视化的更新

利用SARG数据库对ARG进行注释和量化的集成工具被称为ARGs-OAP[†]。它在3.0版本中进行了改进，可以精确地从宏基因组数据集中量化ARG丰度，采用以下修正方程：

$$\text{Abundance} = \sum_{i=1}^{n}\left(k \times \frac{Ni_{\text{ARG-like sequence}} \times Li_{\text{read}} / Li_{\text{ARGs reference sequence}}}{N_{\text{cell number}}}\right)$$

其中，$Ni_{\text{ARG-like sequence}}$是注释到一个特定ARG参考序列的类ARG读取的数量；Li_{read}是读取的长度；$Li_{\text{ARGs reference sequence}}$为对应的ARG参考序列的核苷酸序列长度；$n$为属于该ARG类型或亚型的映射ARG参考序列的数量；$N_{\text{cell number}}$是通过定位到一个必需单拷贝标记基因数据库或从16S rRNA序列的拷贝数中进行校正来估计的细胞数[9]。如果特定ARG参考序列为双组分系统，则参数k为0.5，如果特定ARG参考序列为三组分系统，则参数k为0.33，对于除上述两类外的所有ARG，参数k为1.0。

在SARG v3.0中，不同亚型的ARG亚型被标记为不同的k值，用于调整量化。共有3552个序列被标记为“三组分”系统，其k值为0.33，因为真正的抗性需要该类别中的三个基因作为一组同步出现，并且在没有0.33的调整参数的情况下，将每个组分计为1的旧量化方法将导致三倍的高估。同样地，65个序列被标记为“双组分”系统，k值为0.5，因为这一类中的两个基因都出现会导致抗性；因此，单个事件已通过参数0.5进行调整。这种修改将有助于减少少数ARG类型的定量偏差，包括多药ARG、MLS抗性基因等。k值为1.0的ARG亚型在更新公式的量化过程中不受影响。

此外，在线分析平台已经更新为一个新的文件管理系统，在许多方面促进了更用户友好的在线分析的发展。首先，旧版本的ARGs-OAP需要局部样品预处理。通过更新，用户可以选择上传原始读取的内容（通过网页或FTP），然后只需点击一下，就可以完成ARGs-OAP的量化步骤。其次，更新后的在线流程在使用上述量化后提供了多个下游分析。可视化软件包已经集成到ARGs-OAP和下游分析工具中，用来显示结果，以便更好地进行解释。具体来说，环境样本的短读长可以作为ARGs-OAP分析的查询输入，用于对ARG进行分类，并量化ARG流行率，生成ARG类型、亚型和变异的丰度表。还有一个仪表板，其中包含检测到的ARG的汇总计数和ARG丰度的柱状图（以每个细胞的ARG拷贝数为单位）。

以“地理比较”分析为例，下游分析的工作流程如图3和图4所示。整个下游分析软件包包括：

（1）通过输入来自不同地点的ARG类型样本的丰度表进行地理比较，以将ARG污染水平与来自同一类型栖息地的全球数据进行基准比较；

（2）基于微生物源跟踪（MST）的ARG污染源识别，确定不同来源（包括污水、人类粪便、牲畜粪便、污水处理厂、农田、工业污水处理厂、矿山、自然样品）对感兴趣样本中ARG的贡献比例；

（3）通过参考数据库中不同生态系统集合中的ARG谱，对样本进行排序分析（ordination analysis），得到相似性和差异性；

（4）感兴趣样本中，关于ARG风险的四个等级的概况。

除了对短读数进行宏基因组分析外，基于长读数的ARG注释的应用也越来越多，它要么是由第三代测序生成的[31–32]，要么是将短读长从头组装[33]。通过参考SARG数据库，长读数可以很容易地与蛋白质或核苷酸参考序列进行比对[††]，以注释ARG，这取决于测序的准确性和研究场景。通过整合子可视化和识别流程（I-VIP）[34]或质粒分类流程（Plascad）[35]，MGE分析可以进一步解读遗传背景。ARG和MGE共定位的鉴定为进一步探索细菌群落中潜在的水平基因转移提供了关键信息。

3.4. ARGs-OAP v3.0性能评估

我们通过在读取长度分别为150 bp、201 bp和300 bp的模拟宏基因组数据集中注释ARG，根据MCC、灵敏度和精度评估了更新流程的性能（图5、附录A中的图S1）。评估结果显示，ARGs-OAP v3.0具有良好的性能，在三组环境宏基因组中应用推荐的截止点（即E值：1×10^{-7}；相似度：80%；比对长度比：75%），ARGs-OAP v3.0在ARG识别方面表现出很高的精度和灵敏度。在注释复杂样本中的基因时，误报始终是一个问题。事实证明，ARGs-OAP v3.0是高质量的，误报率低于2%。

为了进一步评估其性能，我们应用ARGs-OAP的三个版本（v1.0、v2.2和v3.0）分析了7个典型环境源的36个宏基因组数据集，它们代表了不同程度的人类活动。

[†] https://smile.hku.hk/SARGs.

[††] https://smile.hku.hk/ARGs/Indexing/download.

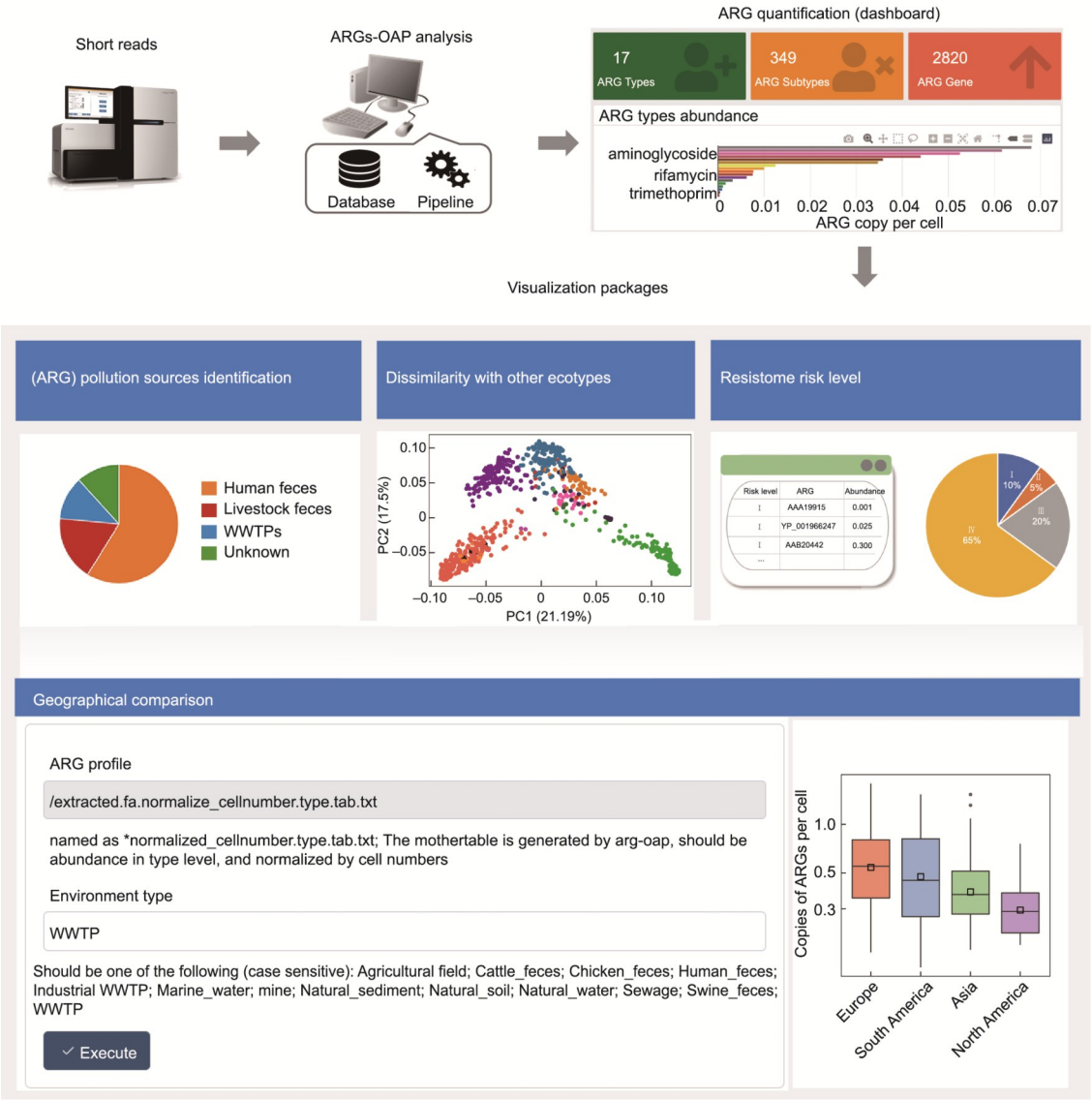


图3. ARGs-OAP v3.0平台的短读长工作流程。查询数据集可以进行分析，以一种高效和准确的方式量化ARG，然后使用集成的工具进行可视化和解释。一个例子是“地理比较”包，它的接口需要选择查询样本的环境类型，然后上传一个输入文件，这是对ARGs-OAP进行分析后的ARG丰度的母表。生成的概要文件包括一个箱形图及一个基于查询样本和存档数据库生成的地图，该数据库包含来自13种栖息地的1427个样本。

结果显示，使用更新的数据库有明显的改善，即在所有研究的不同ARG水平的环境样本中，都发现了ARG丰度（每个细胞的ARG拷贝数）和丰富度（检测到的ARG亚型数量）的增加。

如图5（d）所示，与SARG v2.2相比，应用SARG v3.0使ARG检测度在环境中发生了不同程度的变化，在污水中变化了12.6%，在河水中变化了28.8%。自然环境（河水和沉积物）样品有不同程度的改变（4.7%～28.8%），而来自污水处理系统的样品（WWTP ADS、AS和出水）检测总丰度提高的程度很相似（12.4%～15.9%），其他样品的总丰度有所下降。基于Mann-Whitney测试（$P < 0.05$），发现不同的环境类型中ARG总丰度存在显著差异，进而形成四个丰度水平分层，丰度从高到低如下：牲畜粪便>污水>污水处理厂>自然环境样品。无论应用哪个版本的数据库，ARG丰度水平的分层都保持不变。

自然环境样品和污水处理厂中ARG检出率增加，与此同时，这些样本中的检测丰富度（检测ARG亚型的数量）也有所增加，而污水和牲畜粪便中ARG总丰度的减少归因于在数据库中删除了模糊的参考序列，其主要是多药耐药基因。总的来说，SARG v2.2检测到736个亚型（数据库中有1244个亚型），而SARG v3.0检测到1019个亚型。也就是说，更新后的数据库多检索到了283种ARG亚型，这些亚型来自7种环境类型的至少一个样本中，并

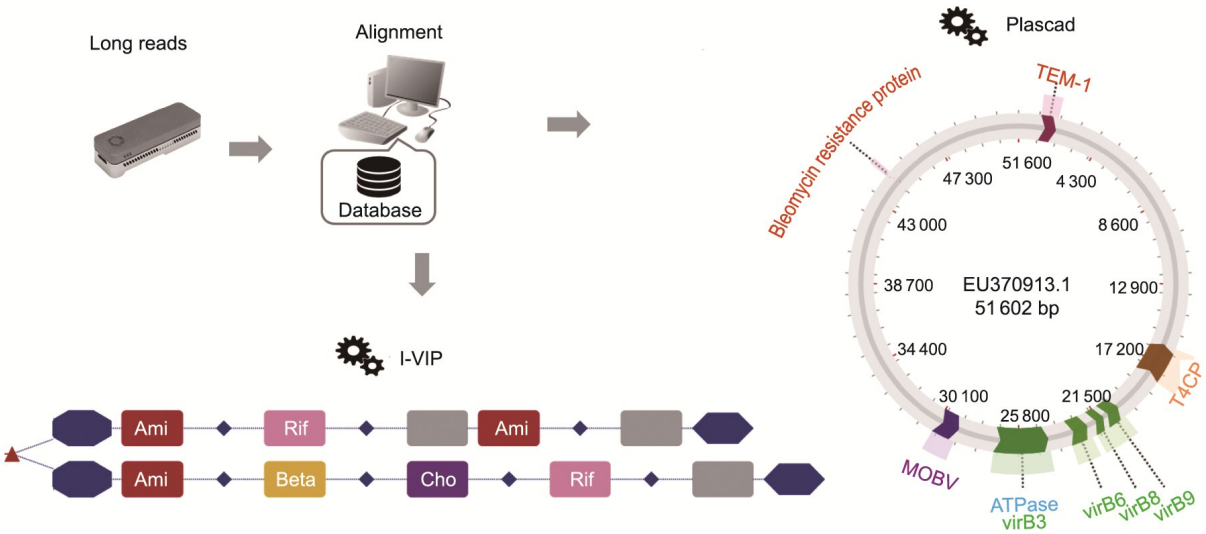


图4. ARGs-OAP v3.0平台上可用于长读长的工具，包括整合子识别和质粒分类。I-VIP：整合子可视化和识别流程。

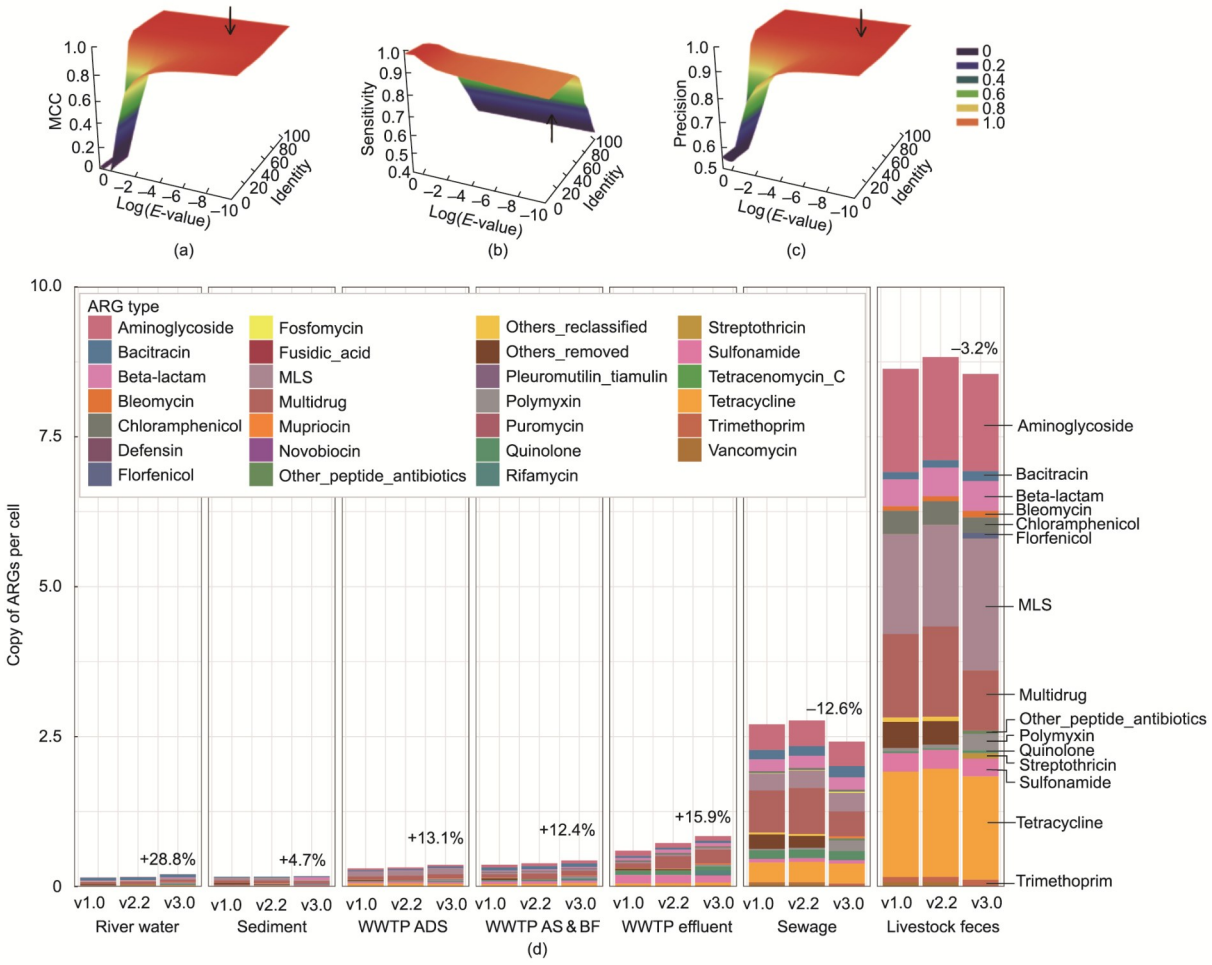


图5. 评估更新后的数据库及用于ARG注释和量化的流程。在读长为150 bp的模拟宏基因组数据集上应用ARGs-OAP v3.0并采用梯度的截断点时，评估MCC（a）、灵敏度（b）和精度（c）。颜色梯度代表了在0～1范围内的MCC（a）、灵敏度（b）和精度（c）的值。（d）通过应用三个版本的ARGs-OAP，对来自不同环境的宏基因组进行进一步评估。对于每个环境，条形图表示所使用的参考数据库：（左）SARG v1.0；（中）SARG v2.2；（右）SARG v3.0。ARG定量单位为每个原核细胞的平均的ARG拷贝数。图中的百分比标签是与版本v2.2相比，使用SARG v3.0检测到的ARG数量的增加。BF：生物膜。

且使用以前的版本没有被检测到（附录A中的表S3）。这些新检索到的亚型包括氨基糖苷、β-内酰胺、MLS、多药、多黏菌素和其他耐药类型，在测试样品中每个细胞的ARG丰度范围为6.17×10^{-6}~0.92个拷贝。新检测到的每个细胞ARG丰度超过0.4个拷贝的亚型包括耐药型MLS中的*lnuC*和*optrA*，以及耐药型氟芬尼考中的*lnuH*和*fexB*，表明添加新的参考基因至SARG v3.0中后，这些新检测到的亚型将被覆盖。因此，更新后的数据库将提高不同环境样本中ARG监测的检测覆盖率。与此同时，新的数据库将通过减少误报来促进ARG的准确预测。

4. 结论

ARGs-OAP于2016年首次发布，并于2018年进行更新。如本研究所述，这种分析工具已继续开发，以在抗生素抗性的环境方面的研究中取得更好的性能。在ARGs-OAP v3.0中，对数据库更新和不同分析工具的集成都进行了改进。首先，参考数据库SARG已经更新到3.0版本，根据更新的知识删除/添加序列，调整类型和亚型的名称，添加机制组和亚组的信息，并在CARD等其他数据库的基础上扩大覆盖范围。SARG v3.0-S（短读长定量的子数据库）和SARG v3.0-E（快速分析的子数据库）排除了与突变、阻遏物和调节因子相关的基因，已经嵌入了ARGs-OAP v3.0作为参考数据库，而SARG v3.0-F可以通过树结构和字典形式进行可视化。其次，从ARGs-OAP开始，使用集成工具开发了用户友好的工作流程，并进行了后续分析，包括风险等级方案、地理比较、MST和与其他生态系统的相似性/不相似性分析。在分析流程中实现了可视化，这将促进数据解释和有效沟通。

致谢

这项工作得到了中国香港特别行政区研究资助委员会的主题性研究计划拨款（T21-705/20-N）的大力支持。殷晓乐博士要感谢香港大学提供的博士后奖学金。郑夏婉女士感谢香港大学提供的研究生奖学金。作者还感谢实验室技术员Vicky冯女士在整个实验过程中的协助。计算是使用香港大学资讯科技服务中心提供的研究计算设施进行的。

Compliance with ethics guidelines

Xiaole Yin, Xiawan Zheng, Liguan Li, An-Ni Zhang, Xiao-Tao Jiang, and Tong Zhang declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary material to this article can be found online at https://doi.org/10.1016/j.eng.2022.10.011.

References

[1] Danko D, Bezdan D, Afshin EE, Ahsanuddin S, Bhattacharya C, Butler DJ, et al. A global metagenomic map of urban microbiomes and antimicrobial resistance. Cell 2021;184(13):3376–93.
[2] Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, Lukjancenko O, et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat Commun 2019;10(1):1124.
[3] Boolchandani M, D'Souza AW, Dantas G. Sequencing-based methods and resources to study antimicrobial resistance. Nat Rev Genet 2019;20(6):356–70.
[4] McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 2013;57(7):3348–57.
[5] Liu B, Pop M. ARDB-antibiotic resistance genes database. Nucleic Acids Res 2009;37(Suppl 1):D443–7.
[6] Yang Y, Jiang X, Chai B, Ma L, Li B, Zhang A, et al. ARGs-OAP: online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database. Bioinformatics 2016;32(15):2346–51.
[7] Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26(19):2460–1.
[8] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403–10.
[9] Yin X, Jiang XT, Chai B, Li L, Yang Y, Cole JR, et al. ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. Bioinformatics 2018;34(13):2263–70.
[10] Roberts MC, Schwarz S. Tetracycline and chloramphenicol resistance mechanisms. In: Mayers DL, Sobel JD, Ouellette M, Kaye KS, Marchaim D, editors. Antimicrobial drug resistance: mechanisms of drug resistance. New York City: Springer; 2017. p. 231–43.
[11] Chopra I, Roberts M. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol Mol Biol Rev 2001;65(2):232–60.
[12] Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 2017;45(D1):D566–73.
[13] Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 2000;28(1):45–8.
[14] Zhang AN, Gaston JM, Dai CL, Zhao S, Poyet M, Groussin M, et al. An omicsbased framework for assessing the health risk of antimicrobial resistance genes. Nat Commun 2021;12(1):4765.
[15] Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 2018;46(W1):W537–44.
[16] Roberts MC. Update on macrolide-lincosamide-streptogramin, ketolide, and oxazolidinone resistance genes. FEMS Microbiol Lett 2008;282(2):147–59.
[17] De Oliveira DMP, Forde BM, Kidd TJ, Harris PNA, Schembri MA, Beatson SA, et al. Antimicrobial resistance in ESKAPE pathogens. Clin Microbiol Rev 2020;33(3):e00181–219.
[18] Munita JM, Arias CA. Mechanisms of antibiotic resistance. Microbiol Spectr 2016;4(2):VMBF-0016–2015.
[19] Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJV. Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol 2015;13(1):42–51.
[20] Bush K. The ABCD's of b-lactamase nomenclature. J Infect Chemother 2013;19(4):549–59.
[21] Rodríguez-Martínez JM, Velasco C, Álvaro P, Cano ME, Luis MM. Plasmid-

mediated quinolone resistance: an update. J Infect Chemother 2011; 17(2): 149–82.
[22] Wright GD. Q&A: antibiotic resistance: where does it come from and what can we do about it? BMC Biol 2010;8(1):123.
[23] Ramirez MS, Tolmasky ME. Aminoglycoside modifying enzymes. Drug Resist Updat 2010;13(6):151–71.
[24] Piddock LJV. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. Clin Microbiol Rev 2006;19(2):382–402.
[25] Poole K. Efflux-mediated antimicrobial resistance. J Antimicrob Chemother 2005;56(1):20–51.
[26] Connell SR, Tracz DM, Nierhaus KH, Taylor DE. Ribosomal protection proteins and their mechanism of tetracycline resistance. Antimicrob Agents Chemother 2003;47(12):3675–81.
[27] Warburton PJ, Ciric L, Lerner A, Seville LA, Roberts AP, Mullany P, et al. TetAB46, a predicted heterodimeric ABC transporter conferring tetracycline resistance in *Streptococcus australis* isolated from the oral cavity. J Antimicrob Chemother 2013;68(1):17–22.
[28] Nishino K, Yamada J, Hirakawa H, Hirata T, Yamaguchi A. Roles of TolC-dependent multidrug transporters of *Escherichia coli* in resistance to β-lactams. Antimicrob Agents Chemother 2003;47(9):3030–3.
[29] The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 2021;49(D1):D480–9.
[30] Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res 2021;49(D1):D412–9.
[31] Yang Y, Zhang AN, Che Y, Liu L, Deng Y, Zhang T. Underrepresented high diversity of class 1 integrons in the environment uncovered by PacBio sequencing using a new primer. Sci Total Environ 2021;787:147611.
[32] Che Y, Xia Y, Liu L, Li AD, Yang Y, Zhang T. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome 2019;7(1):44.
[33] Ma L, Xia Y, Li B, Yang Y, Li LG, Tiedje JM, et al. Metagenomic assembly reveals hosts of antibiotic resistance genes and the shared resistome in pig, chicken and human feces. Environ Sci Technol 2016;50(1):420–7.
[34] Zhang AN, Li LG, Ma L, Gillings MR, Tiedje JM, Zhang T. Conserved phylogenetic distribution and limited antibiotic resistance of class 1 integrons revealed by assessing the bacterial genome and plasmid collection. Microbiome 2018;6(1):130.
[35] Che Y, Yang Y, Xu X, Břinda K, Polz MF, Hanage WP, et al. Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. Proc Natl Acad Sci USA 2021; 118(6): e2008731118.