

微生物群落多样性数学表征方法及其在污水处理系统研究中的应用

夏瑜¹ 何绪文^{1*} 文湘华²

(1. 中国矿业大学(北京)化学与环境工程学院 北京 100083)

(2. 清华大学环境学院 环境模拟与污染控制国家重点实验室 北京 100084)

摘要: 微生物群落在调节全球气候、人类健康和工业生物技术应用中扮演着重要角色。定量表征微生物群落多样性是认识微生物群落基本特征、动态变化和功能的前提。本文介绍了常用的 α 多样性指数, 包括物种数目、Shannon-Weaver 指数、Simpson 多样性指数和 Hill 多样性指数; 并介绍了其他多样性评估方法及其原理, 包括能够评价样本对微生物群落各物种覆盖程度的稀释性曲线和 Good's coverage 指数, 以及能估算群落多样性的 Chao1、ACE 指数和基于物种丰度分布曲线的模型方法; 并以最大规模的生物技术应用——污水处理厂为例, 介绍了这些方法在认识微生物群落多样性中的应用。现有研究表明: 所检测到的城市污水处理厂中微生物群落的物种数目和 Shannon-Weaver 指数随检测方法解析通量(样本大小)的增加而增大; 但现有方法仍无法反映城市污水处理厂微生物群落的真实多样性。基于特定的物种丰度分布曲线对 DNA 样本数据进行模拟和重建, 结果表明对群落物种数目的评估存在较大的不确定性; Shannon-Weaver 指数, 特别是 Simpson 多样性指数等受低丰度物种影响较小, 可以准确计算, 是评价和比较微生物群落分类学多样性的好手段。改进模型拟合方法和加大取样深度能提高微生物群落物种数目的评估精度。此外, 认识微生物群落其他方面的多样性如系统发育多样性和功能多样性, 也对认识微生物生态特征具有重要意义。

关键词: 多样性, 微生物群落, α 多样性指数, 稀释性曲线, 物种丰度分布

Foundation item: National Major Science and Technology Project for Water Pollution Control and Treatment (2017ZX07402002)

*Corresponding author: Tel: 86-10-62339885; E-mail: hjinghua@vip.sina.com

Received: October 23, 2017; Accepted: December 19, 2017; Published online (www.cnki.net): January 16, 2018
基金项目: 国家水体污染控制与治理科技重大专项(2017ZX07402002)

*通信作者: Tel: 86-10-62339885; E-mail: hjinghua@vip.sina.com

收稿日期: 2017-10-23; 接受日期: 2017-12-19; 网络首发日期(www.cnki.net): 2018-01-16

Mathematic methods for the evaluation of microbial diversity and their applications in the research on wastewater treatment systems

XIA Yu¹ HE Xu-Wen^{1*} WEN Xiang-Hua²

(1. School of Chemical and Environmental Engineering, China University of Mining & Technology, Beijing 100083, China)

(2. State Key Joint Laboratory of Environmental Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China)

Abstract: Microbial communities play central roles in global climate regulation, human health and industrial biotechnology. The quantification of microbial diversity is important for the understanding of ecological characteristics of communities, their dynamics and functions. Herein, we introduced the commonly used alpha-diversity indices, including richness, Shannon-Weaver index, Simpson diversity indices and Hill's diversity number. Also, we summarized diversity evaluation ways which are used in estimating the coverage of molecular methods (e.g., rarefaction curve and good's coverage) and community richness (e.g., Chao1 and ACE indices and taxa abundance distribution curve). Then we showed the application of mathematic methods in the research on microbial diversity by taking wastewater treatment plant (WWTP) as an example which is the largest application of bioprocess engineering. Current investigations showed that taxa richness and Shannon diversity of activated sludge microbial communities in full-scale WWTPs increased with the increase in sampling sizes of different methods. However, the disparity between sample size and community size is a common problem in microbial investigations. By reconstructing microbial communities using DNA sampling data based on certain taxa abundance distribution curves, diversity of the communities was evaluated. It was shown that microbial richness was characterized with large uncertainty. Shannon diversity, and especially Simpson diversity indices which are weakly dependent on low-abundance taxa could be estimated accurately. They are good tools to evaluate and compare microbial taxonomic diversity. Developing novel modeling approaches and advances in sequencing technology can improve the accuracy of microbial richness evaluation. In addition, clarifying phylogenetic and functional diversity of microbial communities are also of substantial importance for the understanding of microbial ecology.

Keywords: Diversity, Microbial community, Alpha-diversity index, Rarefaction curve, Taxa abundance distribution

生物多样性(Biological diversity)是指生命形式的多样性,生物所扮演的生态角色以及所包含的基因多样性。由于基因是各级生物系统生物多样性的根本来源(如分子、生物、种群、物种和生态系统),“遗传多样性”和“生物多样性”在一些情况下可以互换使用^[1]。生物多样性是生态群落的重要特征,是群落功能和动态变化的重要影响因素^[2-5]。

微生物群落在调节全球气候、影响人类健康和工业生物技术应用方面发挥着重要作用^[6]。它们常以群体形式存在,具有高度复杂性、非线性、进化与动态变化特征^[7]。以生物技术最大规模的应用——污水处理系统为例,生物处理技术是处理

城市污水和工业废水最经济有效且应用最普遍的方法。微生物群落是污水生物处理系统中碳、氮、磷等污染物去除的主体,认识其多样性是认识污水处理厂的功能和稳定性的重要途径。

微生物具有个体微小、数量巨大、多样性高、可培养率极低的特点。地球生物圈存在的微生物数量多达 10^{30} 个^[8]。据估计,1 g 土壤中包含 10^4 – 10^6 种原核微生物(Prokaryotic species)^[9-10],一个城市污水处理厂中微生物的数量多达 10^{18} 个^[11]。然而,自然界中可培养的微生物数量通常小于 1%^[9],活性污泥中微生物的可培养率为 1%–15%^[12-13],这使得微生物群落的研究十分困难。对微生物群落

的早期研究主要基于培养法。20 世纪 90 年代分子生物学技术的发展与应用为微生物群落研究提供了极大便利,使研究者能在 DNA 分子水平上对微生物特性进行认识,分析范围扩大到数量占绝大多数的不可培养微生物,分析通量也大大增加。特别是高通量测序技术的发展与应用使得研究者对微生物群落的认识到达前所未有的水平。然而在微生物群落研究中,样本大小与微生物群落中实际个体数目仍然相差甚远,远远大于动植物研究中样本与实际群落间的差异,这是微生物群落研究普遍面临的问题^[11]。因此,借助合理的数学方法来认识和推测数量巨大且信息不被样本完全反映的微生物群落的特征具有重要意义。

现有研究通常借助宏观生态学(即动植物生态学)研究中的方法来表征微生物群落的多样性。在生态学研究,有 3 个重要的多样性指数—— α 、 β 、 γ 多样性指数。在微生物群落的分析中,主要涉及到前两者。其中, α 多样性指数主要关注局域均匀生境下的物种数目,因此也被称为生境内的多样性(Within-habitat diversity)。 β 多样性指沿环境梯度不同生境群落之间物种组成的相异性或物种沿环境梯度的更替速率,也被称为生境间的多样性(Between-habitat diversity)。

本文将对目前应用较为广泛的 α 多样性指数的计算公式及其原理进行介绍,并以污水处理系统微生物的研究为例,介绍各指数在微生物群落多样性分析中的应用;随后,针对目前取样大小与真实微生物群落中的个体数目存有较大差异这一现状,总结评估样本对微生物群落多样性覆盖度的方法及其局限性,以及根据测试数据来估算微生物群落多样性的常用方法及其局限性,以期为研究者选用科学的方法定量评估和对比微生物群落多样性提供参考。

1 常用的 α 多样性指数及其应用

目前,常用的 α 多样性指数包括物种数目、Shannon-Weaver 指数、Simpson 多样性指数和 Hill 多样性指数等。它们被应用于污水处理系统活性污

泥微生物群落多样性的表征,也被应用于不同污水处理系统微生物群落多样性的比较、污水处理系统与其他生态环境间的微生物群落多样性的比较、评估环境因素变化对污水处理系统微生物群落多样性的影响等方面。微生物群落 α 多样性指数的计算是以对微生物群落进行物种划分为基础的。然而,目前对微生物物种划分的标准及单位尚无明确定义,且因各类解析方法对微生物的分辨率有所不同,其分类标准也不一。例如,培养法对微生物的分类主要基于其形态特征、生理生化特征和生态学特性等方面的特点。在分子生物学研究中,常根据微生物的基因相似性对微生物进行分类。由变性梯度凝胶电泳(DEEG)方法所得的微生物数量以条带数计;由末端限制性片段长度多态性(T-RFLP)方法所得的微生物数量以限制性片段数量计;在测序数据中,常将相似度 $\geq 97\%$ 的 DNA 微生物序列划分为一个操作分类单位(Operational taxonomic unit, OTU)。因此,在评价和比较微生物群落的多样性时,需要定义微生物物种划分的标准。本文介绍的基于测序数据的活性污泥多样性计算以序列相似度 $\geq 97\%$ 为物种(Taxa)划分标准。

1.1 物种数目及其应用

微生物种类繁多,群落中微生物的物种数目,即丰富度(Richness)是衡量其多样性的一种形式。在对城市污水处理厂好氧活性污泥微生物群落的调查中,早期基于培养法的典型研究实现了对单个活性污泥样品中几十个菌株的解析,检测到的微生物属的数量为个位数^[14-15];基于克隆文库研究的解析通量通常为几十到数百个克隆子^[16-18],检测到的 OTU 数目为几十个^[17-18];基于 16S rRNA 基因的高通量测序研究解析了数千至数十万条序列,检测到了数百个属、数千至一万余个 OTU^[19-23];基于 16S rRNA 基因的寡核苷酸芯片 PhyloChip 对活性污泥微生物群落多样性的研究,检测到的 OTU 数目为数千个^[24]。由此可知,分子生物学方法,特别是高通量技术的应用使得在污水处理系统微生物群落样品中检测到的物种数目呈现数量级的增长(表 1)。

表 1 实际规模的城市污水处理系统微生物群落多样性定量表征结果

Table 1 Quantification of microbial diversity of full-scale municipal wastewater treatment systems

Methods	Sample size	Richness	Shannon-Weaver index	Simpson's index of diversity	Chao1	ACE	Good's coverage
Isolation ^[14]	39–57 strains	5–8 genera	1.06–1.44 a	/	/	/	/
Isolation ^[15]	60 and 69 strains	9 and 11 genera	1.89 and 1.93 a	/	/	/	/
Clone library ^[16]	53–78 isolates	/	3.68–5.12 b	/	/	/	/
Clone library ^[17]	>200 isolates	33 OTUs	3.01 b	/	/	/	/
Clone library ^[18]	119 isolates	94 OTUs	4.42 b	/	/	/	/
DGGE ^[25]	/	7–17 Bands	0.72–1.12 c	/	/	/	/
454 pyrosequencing ^[19]	18 808 sequences	2 455 OTUs	6.42 b	/	3 902 b	/	0.942 b
454 pyrosequencing ^[20]	16 489 sequences	1 183–3 567 OTUs	/	/	2 044–7 086 b	2 146–6 934 b	0.861–0.959 b
454 pyrosequencing ^[21]	7 422–11 151 sequences	2 176–4 123 OTUs	6.26–7.36 b	/	/	/	/
Illumina MiSeq sequencing ^[22]	23 970–46 429 sequences	7 156–10 714 OTUs	7.32–7.99 b	/	31 196–43 581 b	63 106–91 613 b	0.69–0.78 b
Illumina MiSeq sequencing ^[23]	9 923–172 859 sequences	715–1 003 OTUs	5.46–5.98 b	/	1 182–1 776 b	/	/
Phylochip ^[24]	/	1 126–1 729 OTUs	7.0–7.4 b	0.985 9–0.993 6 b	/	/	/

注：a：以属为物种单位计算所得；b：以按序列相似度≥97%的标准划分的 OTU 为物种单位计算所得；c：以 DGGE 条带为物种单位计算所得。

Note: a: Calculated based on genera; b: Calculated based on OTUs (with a 3% nucleotide cutoff); c: Calculated based on DGGE bands.

1.2 Shannon-Weaver 指数及其应用

Shannon-Weaver 指数(H)从两方面来评估微生物群落的多样性：一为物种数目，即丰富度；二为群落中全部物种的个体数目的分配状况，即均匀性(Evenness)。其计算公式为：

$$H = -\sum_{i=1}^S p_i \ln p_i$$

式中， S 表示微生物群落中物种的总数目， p_i 表示物种 i 中包含的所有个体占全部个体中的比例。物种数目越多，多样性越大；同样地，物种之间个体分配均匀性的增加也会使多样性提高。Shannon-Weaver 指数是一种对熵或非冗余度的信息论度量。当在一个系统中若每个个体都属于不同物种时，该系统的冗余度最小，熵最大。Shannon-Weaver 指数是环境微生物研究中常用的多样性指数之一。在城市污水处理厂好氧活性污泥样品中，

基于克隆文库方法所得的微生物群落的 Shannon-Weaver 指数为 3.01–5.12^[16–18]，基于变性梯度凝胶电泳(DGGE)方法所得的值为 0.72–1.12^[25]，基于 16S rRNA 基因的高通量测序方法所得的 Shannon-Weaver 指数为 6.26–7.99^[19–23](表 1)。由此可知，污水处理系统微生物群落的 Shannon-Weaver 指数随着分析方法通量的增长而有所增加。

1.3 Simpson 多样性相关指数及其应用

Simpson 多样性(D)通常指三类相关指数中的任何一种。其中，Simpson 指数(Simpson's index)计算的是在某群落中随机取样的两个个体属于同一物种的概率，其表达式如下：

$$\lambda = \sum_{i=1}^S p_i^2$$

式中， p_i 为属于物种 i 中的个体占全部个体中的比例。

应用较多的为 Simpson 多样性指数(Simpson's index of diversity), 其计算的是在某群落中随机取样的两个个体属于不同物种的概率, 其表达式为 $(1-\lambda)$; 另一种相关的多样性指数为 Simpson 指数的倒数(Simpson's reciprocal index, $1/\lambda$)。两者的值越大, 群落的多样性越高。

与 Shannon-Weaver 指数的应用相比, Simpson 多样性指数在污水处理系统微生物群落多样性的评估中应用较少。采用 454 焦磷酸测序方法, 通过对比城市污水处理厂中缺氧-好氧-厌氧(A²O)系统好氧池、氧化沟(OD)系统好氧段及 A²O-膜生物反应器耦合工艺(MBR 工艺)膜池内的微生物群落的 Simpson 多样性指数和均匀性指数, 例如, A²O 系统的 Simpson 多样性指数和均匀性指数分别为 0.997 3 和 0.889 6, 分别高于 MBR 系统的相应值(0.990 2 和 0.849 6), 得到结论 MBR 系统微生物群落的多样性最低。其可能原因为 MBR 系统的污泥停留时间(SRT)较长和有机负荷率(F/M)较低^[26]。另一基于 454 焦磷酸测序的城市污水处理厂厌氧段的微生物群落的 Simpson 多样性指数为 0.99, 高于工业废水处理厂厌氧段的相应值^[27]。

1.4 Hill 多样性指数及其应用

Hill 多样性指数是一种能够有效描述和量化生物多样性的概念框架, 其表达式为:

$$N_a = \left(\sum_{i=1}^S p_i^a \right)^{\frac{1}{1-a}}$$

各物种相对丰度(p_i)的权重(a)不同, 对应的 Hill 多样性指数的形式不同。当 $a=0$ 时, Hill 多样性指数等于物种数目; 当 $a=1$ 时, Hill 多样性指数 N_1 为 Shannon-Weaver 指数的指数形式; 当 $a=2$ 时, N_2 为 Simpson 指数的倒数形式。随着 a 的增大, Hill 多样性指数的值受低丰度物种的影响逐渐减小, 丰度高的物种对群落多样性的贡献率占主导地位。当 a 趋于无穷大, Hill 多样性指数等于丰度最高物种的相对丰度的倒数形式($1/p_{\max}$), 其被称为 Berger-Parker 指数^[28]。

Hill 多样性指数被用于评估污水处理厂微生物

群落多样性受环境干扰(SRT 历经从 30 d 依次降至 12 d 和 3 d 并最终恢复至 30 d 的变化过程)的影响。通过对比污水处理厂最初及 SRT 经系列变化恢复至 30 d 后的微生物群落多样性指数($a=1$ 和 $a=2$ 时的 Hill 多样性指数, 即 N_1 和 N_2 值), 得到结论: 微生物群落中丰度高的微生物(Common OTU)恢复较快(N_2 值无显著变化), 但干扰导致群落损失了较多低丰度微生物(Rare OTU)(N_1 值减小), 以致微生物群落总多样性减小^[29]。Berger-Parker 指数被应用于不同环境下的微生物多样性的排序, 得到的排序结果为土壤>海水>活性污泥, 这与实际检测结果一致^[30]。

模拟研究表明, 当研究方法对微生物群落中低丰度物种的检测能力增加, 其检测限由 1% 降低至 0.1% 时, Simpson 指数的倒数(N_2)变化不明显, 表明其对群落中低丰度的微生物不敏感^[31]; 此外, 依据物种丰度分布曲线模型——幂律分布通过 DNA 数据来重新构建微生物群落, 得到结论: 依据 DNA 指纹图谱数据, 仅有 Simpson 指数的倒数(N_2)能被精确评估(变化幅度约为 10%)^[32]; 对于高通量测序数据, Shannon-Weaver 和 Simpson 指数($a \geq 1$)而非物种数目能较为准确地被估计, 其被视为评估和对比微生物群落分类学多样性的好工具^[33]。活性污泥微生物的实际研究也得到类似结果: 减少各样品中由 454 焦磷酸测序所得的序列数目, 所得 OTU 数目有所减小, Shannon-Weaver 指数仅出现略微下降, Simpson 多样性指数和均匀性则基本不变。例如, 当某样品中的序列由 21 762 条随机减小至 11 161 条时, OTU 数目由 5 383 个降至 3 191 个, Shannon-Weaver 指数由 6.783 0 小幅下降至 6.380 4, Simpson 多样性指数由 0.983 6 极小幅度地下降至 0.979 3, 均匀度指数则由 0.789 6 变为 0.790 8, 其变化幅度分别为 -40.7%、-5.9%、-0.4% 和 +0.2%^[26]。

Hill 多样性指数提供了一套评估群落多样性的指数。各指数反映了群落的不同性质, 因此, 需基于特定的研究问题选择合适的多样性指数, 并应注重分析方法检测限和样本大小的差异。在进行不同

污水处理系统微生物多样性的对比时,由不同解析方法产生的数据不具可比性。当由同种方法而得各群落样本在大小上存在较大差异时,例如:若各测序样本中所含序列数目差别较大,为避免错误结论,可选用对低丰度物种不敏感的多样性指数比较多样性,如 Shannon-Weaver 指数和 Simpson 多样性指数等;或对各样品进行重采样(Resample)处理从而对样本大小进行统一^[19-20,34-35]。值得注意的是, Hill 多样性指数中 a 值越高,对应的多样性指数对低丰度物种信息的反映程度越小、受高丰度物种信息影响越大。选用多样性指数时需兼顾其对微生物群落信息的反映程度和其不确定性受低丰度物种的影响,尤其是在对群落中低丰度物种信息较为关注的研究中。

2 样本大小合理性的多样性评估方法及其应用

现有研究无法对微生物群落的个体进行一一调查。一些表征微生物群落多样性的方法也可用来评价样本对微生物群落真实多样性的反映程度,例如,稀释性曲线和 Good's coverage 指数。

2.1 稀释性曲线及其应用

从微生物群落的 DNA 序列样本中随机抽取一定数量的个体(序列),以个体数目为 X 轴、这些个体所代表物种数目为 Y 轴构建的曲线为稀释性曲线(Rarefaction curve)。它可以用来比较不同微生物群落中物种的丰富度,也可用于评估样本大小的合理性。稀释性曲线图中,当曲线趋向平坦时,说明取样的数量合理,分析群落中更多序列仅能观察到少量新物种;反之则表明增加测序深度可以检测到较多新物种。

目前,稀释性曲线在城市污水处理厂微生物群落测序研究中得到了少数应用,这些研究表明,即使采用高通量测序方法(26 906 条序列/样品)所得的稀释性曲线仍无法趋于平坦,表明测序分析深度仍然不足^[36];另一研究也得到了相似结论,测序深度为 18 808 条序列/样品时,样本无法覆盖好氧活性

污泥和进、出水的微生物群落多样性。通过各微生物样本的 OTU 数量、Shannon-Weaver 指数发现:相较于进、出水,污水处理厂好氧段的微生物群落多样性指数更高,厌氧消化反应器的更低,稀释性曲线的结果也支持这一结论^[19]。此外,稀释性曲线的应用表明城市污水处理厂的微生物群落多样性高于工业废水处理厂的群落多样性^[27],也高于试验规模的城市污水处理系统的多样性^[36]。

根据理论分析和计算:当样品大小为 M 时,以 $1/(50M)$ 为标准,定义丰度小于 $1/(50M)$ 的物种为低丰度物种。在不改变低丰度物种总丰度的情况下,改变它们的数量不会对稀释曲线的结果产生明显的影响^[33]。现有的解析数据往往无法提供微生物群落中低丰度物种数目的准确信息,因而对低丰度物种敏感度较低的稀释性曲线可能无法提供准确的群落多样性信息。某一模拟结果表明:在群落 1 的总物种数目不足群落 3 物种数目的 $1/20$ 、其非低丰度物种的数量比群落 3 多 15 倍的情况下,稀释性曲线会呈现错误结论——群落 1 的总多样性值高于群落 3^[33]。因此,应谨慎应用稀释性曲线,尤其是在对比低丰度物种数量比例相差较大的微生物群落多样性时。

2.2 Good's coverage 及其应用

Good's coverage 表示所有不是单个体物种在总样本中的比例,具体计算公式为:

$$C = 1 - (F_1/N)$$

其中, F_1 表示单个体物种的数目, N 表示样本中所有物种的总数。Good's coverage 是一种表征微生物群落多样性的方法。在测序数据量一样的情况下, C 的数值越大,样本的物种丰富度越小。随着测序深度的增加,理论上如果不再出现单个体物种,表示已经测到样本中所有物种。所以通过单个体物种在样本中的比值能够简单判定测序是否饱和,因此 Good's coverage 是一个间接判断测序数据是否足够的指标。目前 Good's coverage 在评估污水处理系统微生物群落多样性的应用中较少,且对其评价标准尚无过多报道。城市污水处理厂好氧段的

微生物群落高通量测序数据的 C 值有所差异,一些研究表明该值在 0.86–0.97 之间^[19-20],另有调查报道该值为 0.69–0.78^[22](表 1)。 C 值被广泛应用的前提是微生物群落 Good's coverage 数据得到充分积累,后期有望对其进行区间划分(如高、中、低区间),为判定各方法对微生物群落多样性覆盖度的合理性及优劣提供便利。

3 微生物群落多样性估算方法及其应用

根据样品多样性来推测微生物群落的多样性的方法包括非参数方法(如 Chao1 和 ACE 指数)及模型拟合方法。

3.1 非参数评估方法及其应用

典型非参数评估方法包括 Chao1 和 ACE 指数,用于推荐群落中的物种总数。

Chao1 其典型计算公式和偏差修正公式分别如下:

$$S_{\text{chao1}} = S_{\text{obs}} + \frac{(f_1)^2}{2f_2} \quad (f_2 > 0, \text{典型公式});$$

$$S_{\text{chao1}} = S_{\text{obs}} + \frac{f_1(f_1 - 1)}{2(f_2 + 1)} \quad (\text{偏差修正公式}).$$

式中, S_{chao1} 是估计的物种数, S_{obs} 是检测到的物种数, f_1 是单个体物种的数目, f_2 是只含两个个体的物种数目。

ACE 指数的计算公式为:

$$S_{\text{ACE}} = S_{\text{abund}} + \frac{S_{\text{rare}}}{C_{\text{ACE}}} + \frac{f_1}{C_{\text{ACE}}} \gamma_{\text{ACE}}^2$$

式中, γ_{ACE}^2 为变异系数,计算公式为:

$$\gamma_{\text{ACE}}^2 = \max \left[\frac{S_{\text{rare}}}{C_{\text{ACE}}} \frac{\sum_{k=1}^{10} k(k-1)f_k}{n_{\text{rare}}(n_{\text{rare}} - 1)} - 1, 0 \right]$$

S_{rare} 为样品中稀缺物种(所含个体数不超过 10)的数量,其计算公式为:

$$S_{\text{rare}} = \sum_{k=1}^{10} f_k$$

S_{abund} 为高丰度物种(所含个体数量超过 10)的数量,其计算公式为:

$$S_{\text{abund}} = \sum_{k=11}^{S_{\text{obs}}} f_k \quad (S_{\text{obs}} \text{ 为检测到的物种数目})$$

n_{rare} 为稀缺物种(所含个体数不超过 10)包含的

个体总数,其计算公式为:

$$n_{\text{rare}} = \sum_{k=1}^{10} kf_k$$

当 $f_1 = n_{\text{rare}}$ 时, $C_{\text{ACE}} = 0$, 用 Chao1 指数的偏差修正公式来计算 ACE 的数值。

Chao1 和 ACE 指数被少数地应用于城市污水处理厂活性污泥中微生物群落多样性的估计^[19-20,22-23,37],其结果表明,Chao1 和 ACE 指数高于实际检测到的 OTU 数目(表 1)。Chao1 指数对群落多样性的评估也存在与稀释性曲线的类似问题,即对低丰度物种的敏感度低,可能会对微生物群落多样性进行错误的评估^[33]。

3.2 基于物种丰度分布曲线的模型拟合方法及其应用

物种丰度分布曲线是对生态群落中各物种丰度(所包含的个体数目)的描述。与动植物群落相似,微生物群落也呈现低丰度的物种数量占绝大多数而高丰度物种的数量仅为极少数的规律^[38]。物种丰度分布模型对微生物群落物种数目的评估具有积极意义,其应用的前提是选择合适的物种丰度分布模型。

物种丰度分布曲线有多种形式,常见的包括丰度-物种(Abundance-Taxa)曲线和秩-丰度(Rank-Abundance)曲线等。目前,尚无实验数据能对微生物群落物种丰度分布曲线进行准确判定^[30,33,39]。一些理论分析表明:基于原核生物的高度变化和随机性生长的特点,微生物群落中各物种的丰度分布被推测符合对数正态模型^[30,40],尽管这一模型尚不受环境微生物群落样品的高通量测序数据支持^[35,41]。物种丰度分布模型在活性污泥微生物群落和其他环境微生物群落多样性的评估中应用极少。Curtis 等^[30]率先利用对数正态模型,并假定丰度最低物种的数量(S_{min})和其包含的个体数(N_{min})均为 1,对由原位荧光杂交技术(FISH)解析所得的活性污泥微生物数据进行了拟合,估计 1 mL 的混合液悬浮固体(MLSS)中含有 70 个原核生物物种。随着微生物样本大小的增加和模型拟合方法的改进,对微生物群落物种丰度分布曲线的认识将更为深入,这有助于提高对微生物群落中物种数目的评估精度。

4 结论与展望

常用的 α 多样性指数包括物种数目、Shannon-Weaver 指数、Simpson 多样性指数和 Hill 多样性指数等。其中, 在前 3 种指数中高丰度物种所占权重依次递增, Hill 多样性指数提供了一种评估群落多样性的概念框架。由各检测方法反映的污水处理厂微生物群落多样性(物种数目和 Shannon-Weaver 指数)随其通量的增加而增加。通过高通量扩增子测序(数万至数十万条序列/样品), 在活性污泥微生物中能检测到数千至一万余个 OTU、数百种微生物属。

稀释性曲线、Chao1 和 ACE 指数等表明现有检测方法对城市污水处理厂微生物群落中各物种的覆盖程度仍然不足。基于物种丰度分布曲线模型的微生物群落多样性评估结果表明: 对受低丰度物种影响较大的多样性指数的评估不确定性较大, 如物种数目; 对 Shannon-Weaver 指数、特别是 Simpson 指数的评估较为精确, 其是评价和比较微生物群落分类学多样性的好手段。选用多样性指数时需兼顾其对微生物群落信息的反映程度及其不确定性受低丰度物种的影响。改善模型和加大取样深度有望提高对微生物群落物种数目的评估精度。

上述主要介绍的是表征微生物群落分类学多样性的方法。为了更全面、科学地表征微生物群落的多样性, 除了考虑微生物群落的物种数目及丰度, 也应考虑物种的系统发育信息的亲疏关系, 即系统发育多样性, 一些相关指数如 Faith's Phylogenetic Diversity (PD)和 PD whole tree 等在污水处理系统微生物群落研究中已初见应用^[23,42]。此外, 微生物群落的功能多样性也是其重要特征。少数研究采用一种新型工具——高通量功能基因芯片 GeoChip 对活性污泥微生物群落展开了调查, 检测到了数万个参与碳、氮、磷循环等过程的功能基因, 并采用 α 多样性指数对微生物群落功能基因多样性进行了表征^[43-45]。这为复杂环境微生物群落功能多样性的认识提供了新型途径, 后期数学方法的发展将进一步促进对微生物群落功能多样性的认识。

REFERENCES

- [1] Wilcox BA. *In situ* conservation of genetic resources: determinants of minimum area requirements[A]//Proceedings of the World Congress on National Parks[C]. Washington DC, USA: Smithsonian Institution Press, 1984: 639-647
- [2] Ives AR, Carpenter SR. Stability and diversity of ecosystems[J]. *Science*, 2007, 317(5834): 58-62
- [3] Hai RT, Wang YL, Wang XH, et al. Bacterial community dynamics and taxa-time relationships within two activated sludge bioreactors[J]. *PLoS One*, 2014, 9(3): e90175
- [4] van Elsas JD, Chiurazzi M, Mallon CA, et al. Microbial diversity determines the invasion of soil by a bacterial pathogen[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(4): 1159-1164
- [5] Philippot L, Spor A, Hénault C, et al. Loss in microbial diversity affects nitrogen cycling in soil[J]. *The ISME Journal*, 2013, 7(8): 1609-1619
- [6] Widder S, Allen RJ, Pfeiffer T, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics[J]. *The ISME Journal*, 2016, 10(11): 2557-2568
- [7] Faust K, Raes J. Microbial interactions: from networks to models[J]. *Nature Reviews Microbiology*, 2012, 10(8): 538-550
- [8] Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, 95(12): 6578-6583
- [9] Dykhuizen DE. Santa Rosalia revisited: why are there so many species of bacteria?[J]. *Antonie van Leeuwenhoek*, 1998, 73(1): 25-33
- [10] Gans J, Wolinsky M, Dunbar J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil[J]. *Science*, 2005, 309(5739): 1387-1390
- [11] Woodcock S, Curtis TP, Head IM, et al. Taxa-area relationships for microbes: the unsampled and the unseen[J]. *Ecology Letters*, 2006, 9(7): 805-812
- [12] Wagner M, Amann R, Lemmer H, et al. Probing activated sludge with oligonucleotides specific for proteobacteria: inadequacy of culture-dependent methods for describing microbial community structure[J]. *Applied and Environmental Microbiology*, 1993, 59(5): 1520-1525
- [13] Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation[J]. *Microbiological Reviews*, 1995, 59(1): 143-169
- [14] Dias FF, Bhat JV. Microbial ecology of activated sludge I. Dominant bacteria[J]. *Applied and Environmental Microbiology*, 1964, 12(5): 412-417
- [15] Benedict RG, Carlson DA. Aerobic heterotrophic bacteria in activated sludge[J]. *Water Research*, 1971, 5(11): 1023-1030
- [16] Yang C, Zhang W, Liu RH, et al. Phylogenetic diversity and metabolic potential of activated sludge microbial communities in full-scale wastewater treatment plants[J]. *Environmental Science & Technology*, 2011, 45(17): 7408-7415
- [17] Jin DC, Wang P, Bai ZH, et al. Analysis of bacterial community in bulking sludge using culture-dependent and -independent approaches[J]. *Journal of Environmental Sciences*, 2011, 23(11): 1880-1887

- [18] Xia Y. Microbial communities in municipal wastewater treatment plants of different scales as revealed by clone library[D]. Beijing: Bachelor's Thesis of Beijing University of Chemical Engineering, 2011 (in Chinese)
夏瑜. 基于克隆文库方法的不同规模城市污水处理厂微生物群落研究[D]. 北京: 北京化工大学学士学位论文, 2011
- [19] Ye L, Zhang T. Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing[J]. *Applied Microbiology and Biotechnology*, 2013, 97(6): 2681-2690
- [20] Zhang T, Shao MF, Ye L. 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants[J]. *The ISME Journal*, 2011, 6(6): 1137-1147
- [21] Wang XH, Hu M, Xia Y, et al. Pyrosequencing analysis of bacterial diversity in 14 wastewater treatment systems in China[J]. *Applied and Environmental Microbiology*, 2012, 78(19): 7042-7047
- [22] Gao P, Xu WL, Sontag P, et al. Correlating microbial community compositions with environmental factors in activated sludge from four full-scale municipal wastewater treatment plants in Shanghai, China[J]. *Applied Microbiology and Biotechnology*, 2016, 100(10): 4663-4673
- [23] Liu T, Liu SF, Zheng MS, et al. Performance assessment of full-scale wastewater treatment plants based on seasonal variability of microbial communities via high-throughput sequencing[J]. *PLoS One*, 2016, 11(4): e0152998
- [24] Xia SQ, Duan L, Song YH, et al. Bacterial community structure in geographically distributed biological wastewater treatment reactors[J]. *Environmental Science & Technology*, 2010, 44(19): 7391-7396
- [25] Moura A, Tacão M, Henriques I, et al. Characterization of bacterial diversity in two aerated lagoons of a wastewater treatment plant using PCR-DGGE analysis[J]. *Microbiological Research*, 2009, 164(5): 560-569
- [26] Hu M, Wang XH, Wen XH, et al. Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis[J]. *Bioresource Technology*, 2012, 117: 72-79
- [27] Shu DT, He YL, Yue H, et al. Microbial structures and community functions of anaerobic sludge in six full-scale wastewater treatment plants as revealed by 454 high-throughput pyrosequencing[J]. *Bioresource Technology*, 2015, 186: 163-172
- [28] Berger WH, Parker FL. Diversity of planktonic foraminifera in deep-sea sediments[J]. *Science*, 1970, 168(3937): 1345-1347
- [29] Vuono DC, Benecke J, Henkel J, et al. Disturbance and temporal partitioning of the activated sludge metacommunity[J]. *The ISME Journal*, 2015, 9(2): 425-435
- [30] Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(16): 10494-10499
- [31] Bent SJ, Forney LJ. The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity[J]. *The ISME Journal*, 2008, 2(7): 689-695
- [32] Haegeman B, Sen B, Godon JJ, et al. Only Simpson diversity can be estimated accurately from microbial community fingerprints[J]. *Microbial Ecology*, 2014, 68(2): 169-172
- [33] Haegeman B, Hamelin J, Moriarty J, et al. Robust estimation of microbial diversity in theory and in practice[J]. *The ISME Journal*, 2013, 7(6): 1092-1101
- [34] Ju F, Zhang T. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant[J]. *The ISME Journal*, 2015, 9(3): 683-695
- [35] Xia Y. Diversity and temporal assembly patterns of microbial communities in municipal wastewater treatment systems[D]. Beijing: Doctoral Dissertation of Tsinghua University, 2017 (in Chinese)
夏瑜. 城市污水处理系统微生物多样性和动态变化[D]. 北京: 清华大学博士学位论文, 2017
- [36] Ye L, Shao MF, Zhang T, et al. Analysis of the bacterial community in a laboratory-scale nitrification reactor and a wastewater treatment plant by 454-pyrosequencing[J]. *Water Research*, 2011, 45(15): 4390-4398
- [37] Lee SH, Kang HJ, Park HD. Influence of influent wastewater communities on temporal variation of activated sludge communities[J]. *Water Research*, 2015, 73: 132-144
- [38] Nemergut DR, Schmidt SK, Fukami T, et al. Patterns and processes of microbial community assembly[J]. *Microbiology and Molecular Biology Reviews*, 2013, 77(3): 342-356
- [39] McGill BJ, Etienne RS, Gray JS, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework[J]. *Ecology Letters*, 2007, 10(10): 995-1015
- [40] Engen S, Lande R. Population dynamic models generating the lognormal species abundance distribution[J]. *Mathematical Biosciences*, 1996, 132(2): 169-183
- [41] Kang S, Rodrigues JLM, Ng JP, et al. Hill number as a bacterial diversity measure framework with high-throughput sequence data[J]. *Scientific Reports*, 2016, 6: 38263
- [42] Griffin SJ, Wells GF. Regional synchrony in full-scale activated sludge bioreactors due to deterministic microbial community assembly[J]. *The ISME Journal*, 2017, 11(2): 500-511
- [43] Xia Y, Wang XH, Wen XH, et al. Overall functional gene diversity of microbial communities in three full-scale activated sludge bioreactors[J]. *Applied Microbiology and Biotechnology*, 2014, 98(16): 7233-7242
- [44] Wang XH, Xia Y, Wen XH, et al. Microbial community functional structures in wastewater treatment plants as characterized by GeoChip[J]. *PLoS One*, 2014, 9(3): e93422
- [45] Xia Y, Hu M, Wen XH, et al. Diversity and interactions of microbial functional genes under differing environmental conditions: insights from a membrane bioreactor and an oxidation ditch[J]. *Scientific Reports*, 2016, 6: 18509