



ICNP 与 SeqCode: 组学时代原核生物命名面临的机遇与挑战

张道锋^{1*}, 焦建宇², 李文均^{2,3*}

1 河海大学海洋学院 江苏省海洋生物资源可持续利用工程研究中心, 江苏 南京 210024

2 中山大学生命科学学院 水产动物疫病防控与健康养殖全国重点实验室, 广东 广州 510275

3 中国科学院新疆生态与地理研究所 荒漠与绿洲生态国家重点实验室, 新疆 乌鲁木齐 830011

张道锋, 焦建宇, 李文均. ICNP 与 SeqCode: 组学时代原核生物命名面临的机遇与挑战[J]. 微生物学报, 2024, 64(7): 2151-2171.

ZHANG Daofeng, JIAO Jianyu, LI Wenjun. ICNP and SeqCode: opportunities and challenges of prokaryotic nomenclature in the omics era[J]. Acta Microbiologica Sinica, 2024, 64(7): 2151-2171.

摘要: 自 1936 年细菌学家 Buchanan 负责起草专门的细菌命名法规以来, 国际原核生物命名法规(International Code of Nomenclature of Prokaryotes, ICNP)在不断发展和完善过程中, 积极促进了原核生物分类学及相关学科的发展。随着组学技术在原核生物多样性研究中的应用, 越来越多未培养的细菌和古菌新类群被发现, 却因为 ICNP 要求活的生物材料作为命名模式(nomenclatural type), 而无法获得生效名称(validly published name)。2022 年, 原核生物命名法规从序列数据描述原核生物命名法(Code of Nomenclature of Prokaryotes Described from Sequence Data, SeqCode)正式发布, 以补充 ICNP 在未培养微生物类群命名方面的不足。SeqCode 不希望和 ICNP 产生较大分歧, 并尽可能保留在将来和 ICNP 合并的可能性。然而, 作为两种独立运行的命名法规, 尚不明确 SeqCode 和 ICNP 并存会对学术界产生怎样的影响。本文系统介绍了 ICNP 和 SeqCode 各自的发展历程和主要内容, 分析了二者的优势和局限性, 并呼吁微生物学相关领域的学者共同关注原核生物命名法规并应用于实践, 以期构建更加合理、有效的原核生物名称系统。

关键词: 从序列数据描述原核生物命名法(SeqCode); 国际原核生物命名法规(ICNP); 命名模式; 未培养原核生物; 分类学

资助项目: 国家自然科学基金(31900001, 9225130002, 32370011, 32200007); 江苏省海洋科技创新项目(JSZRHYKJ202209)
This work was supported by the National Natural Science Foundation of China (31900001, 9225130002, 32370011, 32200007) and the Innovation Project for Marine Science and Technology of Jiangsu Province (JSZRHYKJ202209).

*Corresponding authors. E-mail: LI Wenjun, liwenjun3@mail.sysu.edu.cn; ZHANG Daofeng, zdf@hhu.edu.cn

Received: 2024-01-05; Accepted: 2024-04-22; Published online: 2024-04-26

ICNP and SeqCode: opportunities and challenges of prokaryotic nomenclature in the omics era

ZHANG Daofeng^{1*}, JIAO Jianyu², LI Wenjun^{2,3*}

1 Jiangsu Province Engineering Research Center for Marine Bio-resources Sustainable Utilization, College of Oceanography, Hohai University, Nanjing 210024, Jiangsu, China

2 State Key Laboratory of Prevention and Control of Aquatic animal Diseases and Healthy Aquaculture, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, Guangdong, China

3 State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China

Abstract: Since the formation of a committee under the guidance of R. E. Buchanan in 1936, microbiologists have developed and refined a nomenclatural code for bacteriology, the International Code of Nomenclature of Prokaryotes (ICNP). This code has greatly facilitated recent advances in taxonomy and related fields of microbiology. Technical developments in omics have led to a surge in genome-based discoveries of novel uncultured species, many of which play important roles in environmental and medical microbiology. Nevertheless, the ICNP only recognizes cultures as nomenclatural types, thereby preventing the names of the uncultured microorganisms from being validly published. To compensate for the ineffectiveness of the ICNP in creating permanent and stable names of uncultured taxa, the Code of Nomenclature of Prokaryotes Described from Sequence Data (SeqCode) was published in 2022. The SeqCode was formed with the intention of complementing and someday merging with the ICNP. However, as the two nomenclatural codes currently work independently, the consequences of the concurrence of both codes on the scientific community are still not clear. Here, we introduce the histories and principles of the ICNP and SeqCode, summarize their advantages and limitations, and call for scholars to both respect and utilize the two nomenclatural codes of prokaryotes, aiming to facilitate the establishment of a more reasonable and beneficial naming system.

Keywords: Code of Nomenclature of Prokaryotes Described from Sequence Data (SeqCode); International Code of Nomenclature of Prokaryotes (ICNP); nomenclatural type; uncultured prokaryote; taxonomy

命名法(nomenclature)是与生物实体命名(naming)相关的基本规则、原理和注意事项的综合考量,其目标是在特定的分类系统中促成实体和名称(name)的一一对应。《论语·子路》中提到:“名不正,则言不顺;言不顺,则事不成;事不成,则礼乐不兴”。清晰、明确、合理的命名可以有效促进人们进行交流、记载和认识生物个体等相关活动。好的名称不仅是一个实体

的标识,而且应承载容易被感知的生物学信息。目前,原核生物的命名主要是依据国际原核生物命名法规(International Code of Nomenclature of Prokaryotes, ICNP), ICNP 由国际原核生物系统学委员会(International Committee on Systematics of Prokaryotes, ICSP)进行指导、制定和出版^[1]。ICSP 的司法委员会享有 ICNP 的修订权和原核生物命名事项的解释权,并指导 *International*

*Journal of Systematic and Evolutionary Microbiology (IJSEM)*的出版。

21世纪以来,高通量测序技术已越来越多地应用于微生物群落研究。在不进行纯培养的情况下,采用宏基因组学(metagenomics)和单细胞基因组学技术可以直接获得细菌和古菌基因组,即宏基因组组装基因组(metagenome assembled genome, MAG)和单细胞扩增基因组(single amplified genome, SAG)^[2]。然而,由于大多数未培养微生物的MAG/SAG仍然是不完整的基因组,并且存在被污染的可能性,其保真度一直被质疑^[3]。随着研究技术和生物信息学方法的改进,如三代测序技术的使用,使得MAG的质量非常接近纯菌株基因组的测序质量^[4]。目前,从各种生态系统中已有成千上万未培养微生物的MAG被报道,大大提高了我们对原核生物多样性的认识,描述未培养的“大多数”也是将来的研究热点之一。如果这些未培养微生物能够被规范命名并获得学术界的认可,被有效描述的原核生物将成倍增加,无疑是原核生物命名的重大机遇。

然而,由于缺少公认的命名法规,导致目前未培养微生物的命名不规范、无共识。另一方面,由于要求使用活的生物材料作为命名模式(nomenclatural type),适用于纯培养微生物的命名法规ICNP无法应用于未培养微生物,这在一定程度上制约着微生物多样性研究的发展^[5]。于是针对基因组序列,包容未培养微生物类群的命名法规从序列数据描述原核生物命名法(Code of Nomenclature of Prokaryotes Described from Sequence Data, SeqCode)在2022年正式发布^[6-7]。虽然SeqCode不希望和ICNP产生较大分歧,并尽可能保留在将来和ICNP合并的可能性,但是作为两种独立运行的命名法规,SeqCode和ICNP并存会对学术界产生怎样的影响,仍有

待接受实践检验。因此,原核生物的命名体系和名称系统也面临着严峻挑战。

1 传统的 ICNP 命名法规

1.1 ICNP 发展历程

早在19世纪和20世纪上半叶,细菌学家主要遵循《植物学命名法规》(Botanical Code of Nomenclature)的规则进行命名,因为那时细菌被认为是一类真菌,都属于植物。后来,人们发现植物和细菌的研究方法大不相同,细菌学更注重其培养特征,并认为被命名的模式(type)材料应是活的培养物,而《植物学命名法规》并不强制要求活的模式材料。1930年,在巴黎举行的第一届国际微生物学大会上,有学者建议为细菌制定专门的命名法规。1936年,在伦敦举行的第二届国际微生物学大会上美国细菌学家Buchanan被任命成立一个委员会,负责起草一份细菌命名法规草案(图1),该委员会即后来在1939年正式成立的国际细菌命名委员会(International Committee for Bacteriological Nomenclature, ICBN)^[8]。随后,ICBN在1970年更名为国际细菌系统学委员会(International Committee on Systematic Bacteriology, ICSB),于1999年更名为ICSP,并沿用至今。1936年起草的命名法规经过多次修订后,于1947年编纂完成,定名为《国际细菌生物学命名法规》(International Bacteriological Code of Nomenclature, IBCN),并于1948年正式发布。

1958年,《国际细菌与病毒命名法规》(International Code of Nomenclature of Bacteria and Viruses)正式发布(图1),被认为是现行ICNP的第一版^[8],但也有学者认为1948年版才是第一版^[9]。随着与病毒命名相关的部分独立发展,1966年《国际细菌命名法规》(International Code of Nomenclature of Bacteria, ICNB)正式发

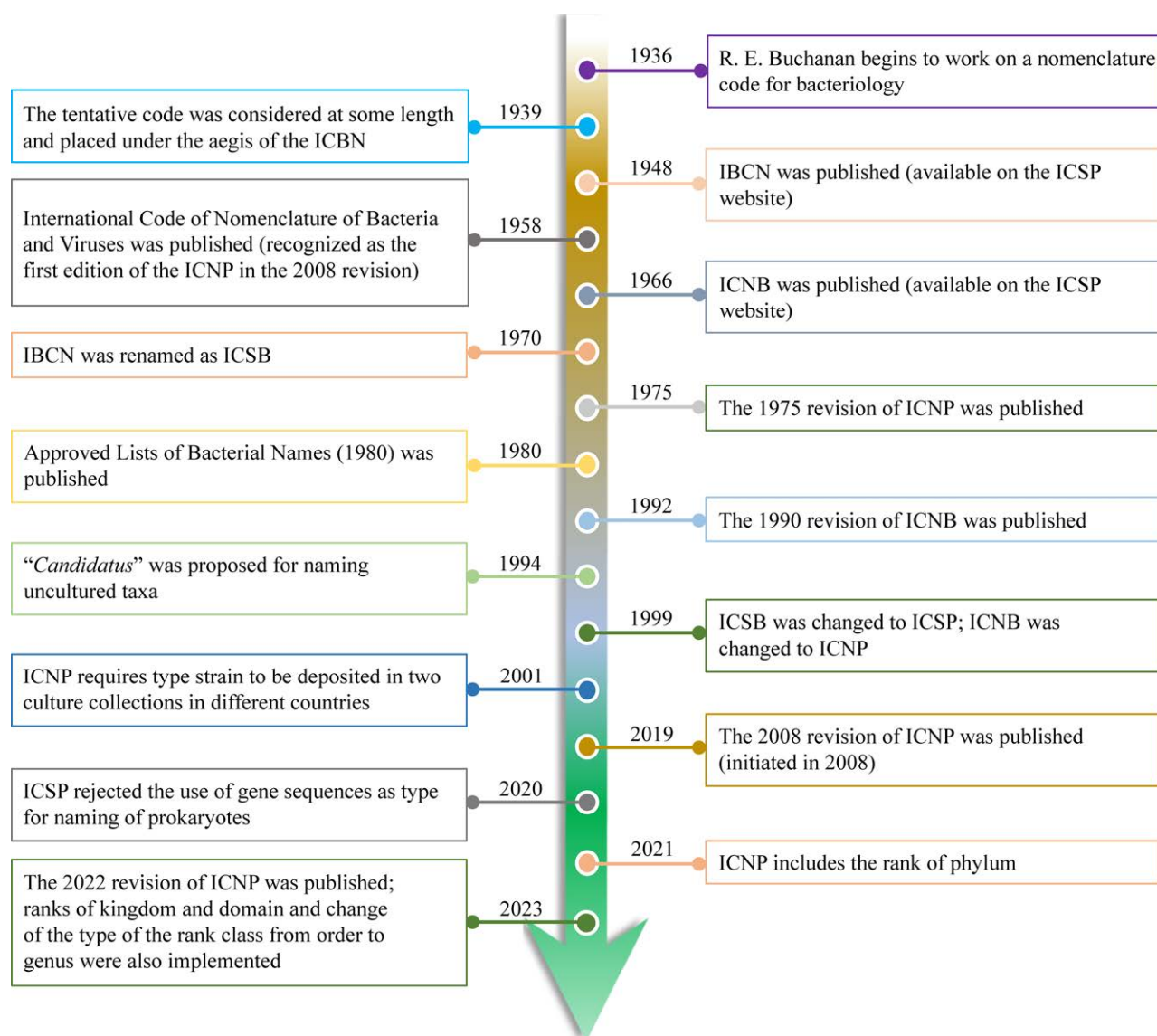


图 1 与国际原核生物命名法规(ICNP)相关的大事记

Figure 1 A brief history of the International Code of Nomenclature of Prokaryotes (ICNP). ICBN: International Committee for Bacteriological Nomenclature. IBCN: International Bacteriological Code of Nomenclature; ICSP: International Committee on Systematics of Prokaryotes; ICSB: International Committee on Systematic Bacteriology; ICNB: International Code of Nomenclature of Bacteria; ICNP: International Code of Nomenclature of Prokaryotes.

布, 该版本尝试简化命名规则, 从而鼓励更多的学者使用该法规。然而该版法规影响力较小, 后来很少被提及。Buchanan 负责的命名法规的最后一次修订发生在 1975 年, 该版本提出了 3 项重要的变革, 至今仍然适用:

建议对已有的细菌名称进行认定并发布认证名录(Approved Lists); 新的名称必须发表在 *International Journal of Systematic Bacteriology* (*IJSB*, *IJSEM* 的前身); 必须为生效发表的名称指定模式材料。

在 1960 年至 1980 年间, 由于缺乏有效的物种描述和模式培养物, 导致大量已发表的细菌名称无法正常使用, 严重影响了微生物学的发展。1980 年, 在澳大利亚细菌学家 Skerman 的号召下, 一个全新的细菌命名法规正式启动。同年, 细菌名称认证名录(Approved Lists of Bacterial Names, 1980)正式发布, 未被该名录收录的细菌名称将被视为未生效发表(图 1)。该名录的发布也是细菌名称系统和分类学走向规范和统一的标志性事件, 被认为是原核生物命名法规诞生以来的第一次重大变革, 对分类学也影响深远^[10]。与此同时, 英国细菌学家 Lapage 牵头重新编写了细菌命名法规, 使其更合乎逻辑, 更便于遵循^[8]。该法规主要包括 2 种类别的内容, 即强制性的规则(rules)和非强制性的建议(recommendations); 对如何描述、命名和发表新型细菌, 以及如何处理命名问题也进行了详细描述。该法规经过多次修订, 在 1992 年以 ICNB 1990 年修订版的形式出版, 并于 1999 年被更名为 ICNP。后一版的 ICNP 于 2008 年开始修订, 于 2014 年第十四届细菌学和应用微生物学(Bacteriology and Applied Microbiology, BAM)大会上发布草案, 并于 2019 年 1 月在 *IJSEM* 上发表。2021 年 2 月, 经 ICSP 委员投票表决后, ICNP 正式将“门”纳入命名系统, 并规定使用“-ota”作为名称后缀^[11]。这直接导致了在此之前长期使用的“门(phylum)”的名称发生了剧烈变化^[12], 在学术界引起了广泛的讨论和争议。现行的 ICNP 2022 年修订版于 2023 年 5 月在 *IJSEM* 杂志上正式发表^[13]。2023 年 8–9 月, ICSP 通过投票表决(34 名委员参与, 24 人赞成)正式将“域(domain)”和“界(kingdom)”纳入 ICNP 命名系统(图 1)^[14]。

1.2 命名法规的基本原则

命名是所有科学研究中的重要组成部分。

如果不能分配精准、明确的名称, 就不可能有效地进行沟通。例如, 在 ICNP 实施和 1980 年的细菌名称认证名录发布之前, 结核分枝杆菌至少有 9 个名称, 存在严重的同物异名。多个名称不仅使查阅文献和交流变得困难, 而且不同名称的确切含义也不尽一致, 这严重阻碍了医学/兽医等领域微生物学的发展。规范统一的命名对于大型数据库的构建和海量数据运算也至关重要, 如果没有一个相对稳定、统一的名称, 数据库的管理会随着时间的推移, 变得越来越困难。在当前生物学数据剧烈增长的背景下, 更需要一个明确的命名法规, 以保障现代生物信息学工具的高效应用。

在生物学中, 大多数命名法规要求生物实体与明确的名称一一对应^[15], 并以“general considerations”和“principles”的形式呈现内容, 以确保每个生物实体在分类系统中对应一个分类地位和唯一的正式名称。在实践中实现这一目标非常困难, 因为命名与生物多样性的发现一直在进行, 而且世界各地的研究人员通常是同时在独立开展这些活动。研究人员通常持有不同的分类学观点和命名理论, 势必导致所创建的名称存在差异。命名法规便致力于促进形成一个统一、稳定、方便理解的命名系统。在这个过程中, 命名法规必须处理好以下 5 个问题^[9]。

1.2.1 生物实体被命名的必要性

对于分类学来讲, 生物实体被称为分类单元(taxon)。与其他命名法规一样, ICNP 对生物实体的定义并不明确, 只是笼统地描述为“any group of organisms treated as a named group in a taxonomy” [consideration 7(3)]。分类学(taxonomy)包括分类(classification)、鉴定(identification)和命名法(nomenclature)三大要素。命名法规聚焦于命名法而不涉及分类和鉴

定，它虽然规定了如何为不同的分类单元(科、属、种等)命名，但如何限定这些分类单元并不在命名法规的职责范围内。因此，ICNP 这种模糊的定义可以确保持有不同分类学观点的研究人员都能够使用该法规。

1.2.2 分类单元存在的证据

在大多数命名法规中，这一证据被称作命名模式(nomenclatural type 或 type)。模式具有 3 个方面的功能：首先，证明这个分类单元的存在；其次，为判定一个新样本是否与该模式属于同一个分类单元提供参考标准；如果代表新的分类单元，这一样本便可以成为“模式”并代表为其命名的名称所指代的分类单元，这便是模式的第 3 个功能。分类单元是以它的模式存在且被命名为基础的，如果一个样本与已知的模式均不相

同，便值得被命名，模式代表了名称的精确含义，是名称所代表生物体的客观体现。

模式主要有两种形式。种和亚种的模式是可用于采集实验数据的生物实体。原核生物的模式便是活着的菌株(图 2)；在 2001 年 1 月 1 日之前，模式也可以是有详细描述、保存完好的失活样本或者插图(图 1)^[16-17]。属及以上高级分类单元的模式是其下属的一个分类单元，例如属有一个模式种，科及以上分类单元均有一个模式属(2023 年以前，并不是所有高级分类单元的模式都是属)。因此，每一个分类单元都可以直接或者间接与证明其存在的实验数据(基于模式种或者亚种)联系在一起。每个名称的确切含义便来源于对分类单元所包含特定模式的认知。

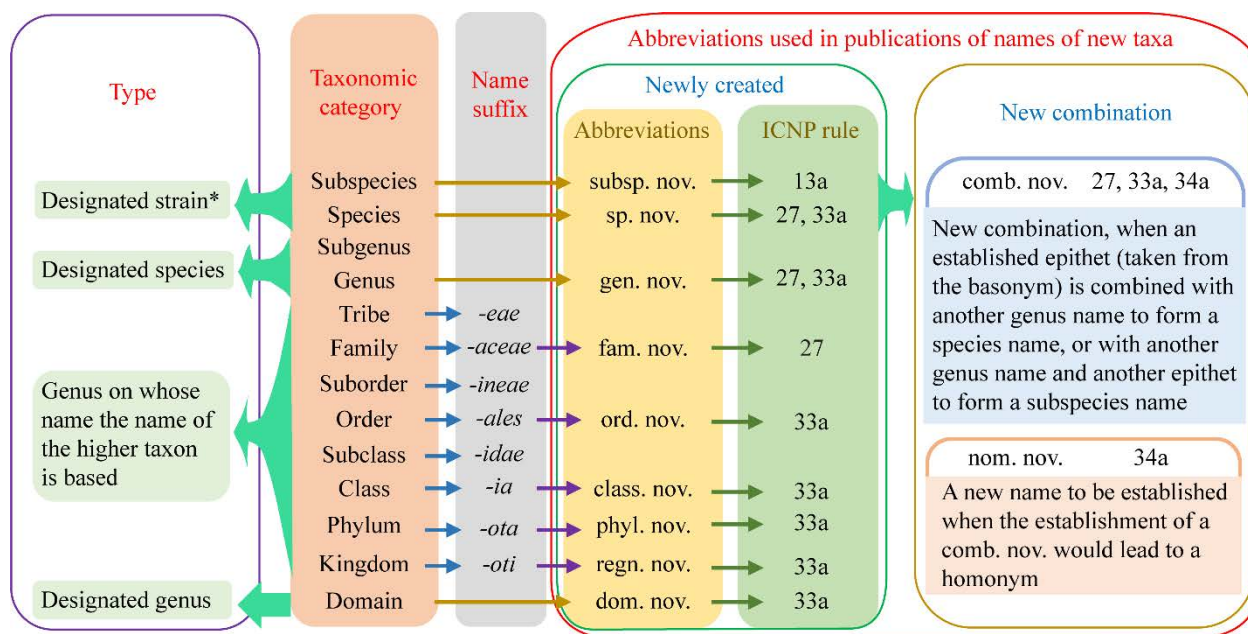


图 2 ICNP 的分类级层、命名规则和模式^[13-14]

Figure 2 Taxonomic ranks and their name suffixes and types in ICNP^[13-14]. New creation means creating names for newly proposed taxa. New combination means creating names (renaming) for existing taxa. *: In special cases the place of the type strain may be taken by a description, preserved specimen, or an illustration (see Rule 18a(1)). Another two abbreviations, “corrig.” and “emend.”, are also frequently seen, “corrig.” indicates a corrected typographical or orthographic error, and “emend.” indicates alteration of the diagnostic characters or of the circumscription of a taxon.

1.2.3 名称

一个好的名称应该便于理解、识别和区分,并具有“顾名思义”功能。自 1751 年起,瑞典博物学家林奈(Linnaeus)提出的以拉丁语和希腊语为词源的“双名法”(属名+种加词,也称为双名命名法)被广泛应用于各类生物的命名,并为整个生物学领域提供了一个稳定、清晰、影响深远的名称系统。然而,从应用的角度,各种来源、形式的名称都是可以接受的^[18],例如用于未培养类群的“字母+数字”的形式在过去的 20 年间就被广泛使用。拉丁语的变化极其多样,同一个词根在不同语境、词性中的书写形式各不相同。古希腊语有自己独特的字母表,不经过专门的学习很难辨认,用于生物命名时通常要先转换成罗马字母(也称拉丁字母)实现拉丁语化才能使用。目前,拉丁语化的名称也仅在欧洲的部分地区才易于理解^[15]。大多数微生物学家都对拉丁语和古希腊语知之甚少,无法创建出符合拉丁语语法规则、结构良好的双名法名称,也无法给出清晰、准确的词源说明。虽然学者们出版了一些指南^[19-22],但是收效甚微。然而,ICNP 坚持使用拉丁语和古希腊语进行命名,并在这方面制定了一些不易理解的规则和建议^[13]。在实践中,拉丁化的名称通常需要相关专家花费大量时间逐一进行质量控制,因而缺乏灵活性。这些专家需要熟悉古典语言、ICNP 和当代微生物学^[20,23],而符合要求的专家已越来越少^[24]。

目前,只有《国际病毒分类与命名法规》(International Code of Virus Classification and Nomenclature)认可非拉丁语的名称,其他一些命名法规也适当放宽了命名规则,不再要求严格遵守拉丁语。使用拉丁语和古希腊语进行命名的优势是可以保证名称体系的延续性和稳定性。然而,在拉丁语和古希腊语日趋“消亡”^[20,24]的背

景下,它们是否能够在实践中继续确保命名系统的稳定性和优雅性,有待商榷。

1.2.4 维护名称系统的稳定性和扩展性

命名法规需要提供一套面对复杂问题时的解决方案,例如在没有统一分类标准的情况下,命名法规仍然能够通过系统地整合大量结果促成统一的名称系统。这套方案需要允许在新分类单元被发现时,新的名称可以被收录;当已有分类单元的分类地位变更时,其名称可以改变且改变过程可追溯;所有组合、变更和拆分的过程必须易于理解和便于采纳;当系统内相互矛盾时,能提供有效的解决方案。在处理这些问题过程中,优先权(priority)是基本原则,即最早生效发表的名称便是相应分类单元的正确名称(correct name)。优先权(即更早的命名时间)确保了名称的稳定性和分类单元名称的唯一性。在种和亚种级别,名称直接与命名模式相关联,因此最早偶联特定模式的名称将对任何后续名称享有优先权。例如,如果一个种被再分类到另一个属,即使属名改变了,种加词也保持不变(如果种加词在新的属内已被占用,尽量在保留原有词根的基础上重新命名)。如果 2 个物种合并,最早生效发表的名称即正确名称。如果一个物种被分割,保留模式的分类单元必须保留原来的名称。因为目前的命名格式为双名法,属名和种加词必须同时被命名,并享有同样的优先权。如果两个属合并,最早被命名的模式种即为新属的模式种,并保留相应种属名。如果一个属被拆分,保留原先模式种的属必须保留原来的属名。

1.2.5 与分类学的关系

大多数命名法规都尽量在分类学问题上保持中立^[23],例如 ICNP 的 Principle 1(4)规定:“nothing in this code may be construed to restrict the freedom of taxonomic thought or action”^[13]。

然而,在实践中,命名法规很难做到绝对中立^[25]。例如,ICNP 和下文要介绍的 SeqCode 都被设计为层次结构的分类名称系统,这是建立在假设层次结构能够合理地反映生物多样性自然属性的基础上。同样地,ICNP 也假设在原核生物中存在被称为“物种”的实体;并假设每一个种都属于一个属,每一个属都隶属于一个更高的等级,即科、目、纲和门。虽然根据人们对现有原核生物多样性的了解,这些基本假设并不完全成立,但是现有命名法规仍然建立了一个非常实用的分类体系,并很好地服务于学术界。

另一方面,命名法规也力求免受分类学其他方面的影响,因为分类(classification)和鉴定(identification)的方法和原则会随着新技术的发展和应用而改变。例如,在 20 世纪 80 年代中期之前,原核生物分类学以“限定性(determinative)”为主要原则,即限定某一分类单元的多样性范围,那时并不强调高级分类单元之间的自然关系。因此,当时的 ICNP 非常重视低级分类单元的实验特性。后来,随着遗传学研究方法的发展,原核生物分类学开始倚重系统发育学方法,导致分类单元的名称发生了大量变化^[26]。在这一过程中,ICNP 便很好地维持了名称系统的稳定性,许多在“限定性”分类学阶段创建的名称在传统分类学向系统发育分类学过渡过程中得以保留。

1.3 ICNP 存在的问题

早在 20 世纪 90 年代,随着 16S rRNA 基因测序等分子生物学技术的发展,对于未获得纯培养物的类群(uncultivated taxa),单纯依靠分子特征对其进行描述成为可能。然而,ICNB 1990 年修订版中并没有与主要依据基因序列对生物实体进行命名相关的规定。一些学者预言,随着相关技术的推广和应用,这一问题会越来越突出,因此向 ICSB 提议将“暂定种(*Candidatus*)”

用于未培养新分类单元的命名^[27]。ICSB 认为:*Candidatus* 名称可以用于命名那些除了序列分析之外未被进一步研究的类群,但是不会被 ICNB 认可为生效名称(validly published under the ICNB);只有相关类群被分离培养和充分鉴定后才能成为生效发表。这是因为 ICNB/ICNP 除了对命名的词源(etymology)进行约束,也对模式的存在证据提出了严格要求(见上文)。ICNP 认为,通过对模式进行严格要求可以有效防止名称滥用和名称指代的生物对象不明确、难观测等现象。例如,从 2001 年开始,ICNP 要求(Rule 30)被命名的模式菌株必须在至少两个国家的两个保藏中心进行保藏。要求模式菌株纯培养物进行保藏的这一规则,决定了现行的 ICNP 不会将基于分子序列发现的海量新分类单元的名称认定为生效发表并赋予其优先权。另外,一些难培养和来自对生物材料转移和共享限制较为严格的国家和地区的菌株,也因为无法满足对模式菌株进行保藏的要求而难以被命名和发表^[5]。

最初,Murray 和 Stackebrandt^[28]关于 *Candidatus* 的建议便是为了缓解 ICNP 中 Rule 30 关于模式菌株保藏的严苛要求。ICNP 2008 修订版的附录 11 (Appendix II)中采纳了这种 *Candidatus* 的命名方式:由 *Candidatus* (斜体,可缩写为 *Ca.*) 后面跟着“属名+种加词”“属名”或“种加词”(均不斜体)组成;描述 *Candidatus* 的学者需要向 *IJSEM* 杂志提交论文的复印件、形式得当的“记录(entry)”和信件说明以申请认定;*IJSEM* 会不定期发布认定名录。然而,ICSP 并未很好地执行这一举措,直到 2020 年 *IJSEM* 才发表第一份 *Candidatus* 认定名录,涵盖了截止到 2018 年底之前的 1 091 个名称^[29-30]。随后,该认定名录以年为区间进行更新^[31-32] (表 1)。然而,*Candidatus* 名称的规范性却存在严重问题:根据 2017 年针

表 1 目前在 *IJSEM* 杂志上发布的 *Candidatus* 名录(引用并修改自文献[31,33])Table 1 The number of '*Candidatus*' names in List No. 1–4 published in *IJSEM* (modified from references [31,33])

Period covered	List No. 1 (2020)	List No. 2 (2021)	List No. 3 (2022)	List No. 4 (2022)
	1995–2018	2019+addenda	2020+addenda	2021+addenda
Classes	7	7+5	9+2	15 +1
Orders	12	11+5	17+5	17+1
Families	25	16+6	19+3	19+1
Genera	329	53+14	46+30	280+4
Species	706	86+24	83+90	1 017+15
Subspecies	10	0	1	0
Other	2	0	0	0
Total	1 091	173+54	175+130	1 348+22
" <i>Candidatus</i> " status lost#	54	2	2	8

Years in brackets mean when the lists were published. Numbers after the plus sign mean addenda names to the previous lists. #: If a '*Candidatus*' taxon was grown in pure culture, and then its *Candidatus*' status lost.

对大约 400 个 *Candidatus* 名称的调查报告, 其中 120 个违背了 ICNP 的命名规则^[34]; 最早的两个 *Candidatus* 认定名录中, 超过 1/4 的名称需要被修正^[30,35]。

另外, “字母+数字(alphanumeric)”的方式也被用于未培养微生物的命名, 尤其是早期被广泛用于指代在系统发育树上聚集成簇的一群 16S rRNA 基因序列, 例如海洋中广泛分布的 SAR11、SAR116 等。相对于 *Candidatus* 名称, “字母+数字”的形式更加缺乏辨识度和统一性, 从而出现了大量同义名称(synonyms, 也称同物异名)^[8,10]。事实上, 最初关于 *Candidatus* 的提议便是希望能够缓解“字母+数字”的命名方式带来的问题。然而, 如上文所述, *Candidatus* 至今未被 ICNP 正式采纳, “字母+数字”的名称带来的问题也一直影响至今。

回顾历史, DNA 序列在物种鉴定中的应用和 ICNP 对 *Candidatus* 的决定影响深远, 被认为是原核生物分类学历史上的第二次重大变革^[10], 这直接导致了未培养类群被大量发现和随意命名, 也为微生物多样性的交流和记载带来了诸

多问题和不便。ICNP 作为原核生物命名法的权威, 却由于自身规则的限制而无法覆盖未培养的“大多数”, 因而在过去 10 年中引发了关于 ICNP 和命名法的激烈讨论。ICNP 仍需要不断更新和完善以覆盖更多的原核生物多样性, 才能更好地服务于学术界。

2 SeqCode 的发展历程

2.1 未培养微生物命名的必要性

随着高通量测序技术越来越多地应用于微生物多样性研究, 人们从各种生态系统中获得了成千上万的 MAG/SAG。这些基因组序列除了被用于物种鉴定之外, 代表高级分类单元的基因组以及种内基因组多样性的发掘也在深刻地改变着细菌和古菌的分类系统。另外, 基因组序列也被用于开发标准化、基于系统发育和序列分化的细菌和古菌分类系统, 例如最具代表性的基因组分类数据库(genome taxonomy database, GTDB; <https://gtdb.ecogenomic.org/>)^[36]。目前(release 09-RS220), GTDB 包含了 11 万多个种的近 59.7 万个基因组, 其中包括大约 28 万个

MAG/SAG,显著提高了人们对细菌和古菌多样性的认识。然而,基于MAG/SAG发现的新分类单元存在命名混乱的问题,例如GTDB使用“字母+数字”的形式,也有很多文献中使用*Candidatus*名称。

与真核生物相比,原核生物被命名和描述的种要少得多。例如,兰科(*Orchidaceae*)植物就有约28 000种,比已知的细菌和古菌都要多。目前已知的生物超过200万种,而微生物只占了其中很小一部分。不管是16S rRNA基因多样性还是MAG/SAG分析均表明,已经获得纯培养的微生物只占自然界的很小一部分。根据GTDB的数据(release 08-RS214):可培养物种的数量不足一半^[23];和正式的名称对应的基因组大约只占1/4^[23];超过65 000个新类群需要被命名^[37]。按照近些年的发展形势,可培养和拥有正式名称物种的比例只会逐渐降低^[23]。目前,描述未培养的“大多数”是研究热点之一,然而未培养类群名称却因为缺少公认的法规而出现形式随意、含义模糊等问题。适用于纯培养微生物的命名法规ICNP又无法应用于未培养微生物(见上文),这在一定程度上制约了微生物多样性研究的发展^[5,7]。

2.2 SeqCode的发展历程

早在2015年,Hedlund等^[38]和Konstantinidis等^[39]就建议修改ICNP中*Candidatus*的概念并将基于MAG和SAG的名称正式纳入其中,以解决未培养类群命名混乱的问题。2016年Whitman^[40]正式在*IJSEM*上发文建议修订ICNP,将高质量的基因组认定为命名模式,这可以使*Candidatus*名称正式生效、确保相关名称的稳定性并形成统一的可培养和未培养微生物命名法规^[41]。2017年,Konstantinidis等^[42]认为未培养微生物需要一个稳定的命名系统,并分析了基于高质量MAG和SAG建立独立命名系统的可能性。

次年,Hedlund和Reysenbach获得了美国国家科学基金会系统和生物多样性科学中心的资助,为下一个10年微生物系统学的发展制定战略。于是,2018年10月和2019年4月分别举行了多次线上和线下研讨会,最终达成了一份共识^[10],提出了两套潜在的解决方案(图3),得到了全球120名知名微生物学家的认可。

方案A(plan A)是通过修订现行的ICNP,在区分可培养和未培养分类单元的情况下,将DNA序列作为未培养类群的命名模式。将从纯培养物(或稳定共培养物)中获得的微生物认定为“可培养”,将通过DNA序列信息识别的微生物认定为“未培养”(包括从混合培养物中获取的DNA序列)。方案A通过在ICSP下设委员会开展工作:审查已有的*Candidatus*名称;制定DNA序列数据标准;规范*Candidatus*名称并统一相关术语的用法(例如分别表示uncultivated、*Candidatus*和environmental的上标u、c和e)以明确区分未培养与可培养物微生物;审核已发表的、以DNA序列作为模式的分类单元并制定认定名录(Approved Lists 2.0)。后续新的名称和相关分类单元的描述(包括词源、类群特性、推测的表型和序列登录号),无论是基于单个还是大规模MAG/SAG的研究,都将被*IJSEM*审核、认定并保护优先权。方案A是在现有的ICNP基础上进行改进,有利于形成一个包含可培养和未培养微生物的统一命名法规。

方案B是为未培养类群制定一套独立于ICNP的命名法规,由国际化的权威组织负责管理。该组织负责制定和维护《国际未培养原核生物命名法规》(International Code of Nomenclature of Uncultivated Prokaryotes, ICNUP;即后来的SeqCode),监督ICNUP的实施,并发布生效的未培养类群名录。ICNUP将任命一个临时委员会来处理已有的*Candidatus*名称,制定生效名

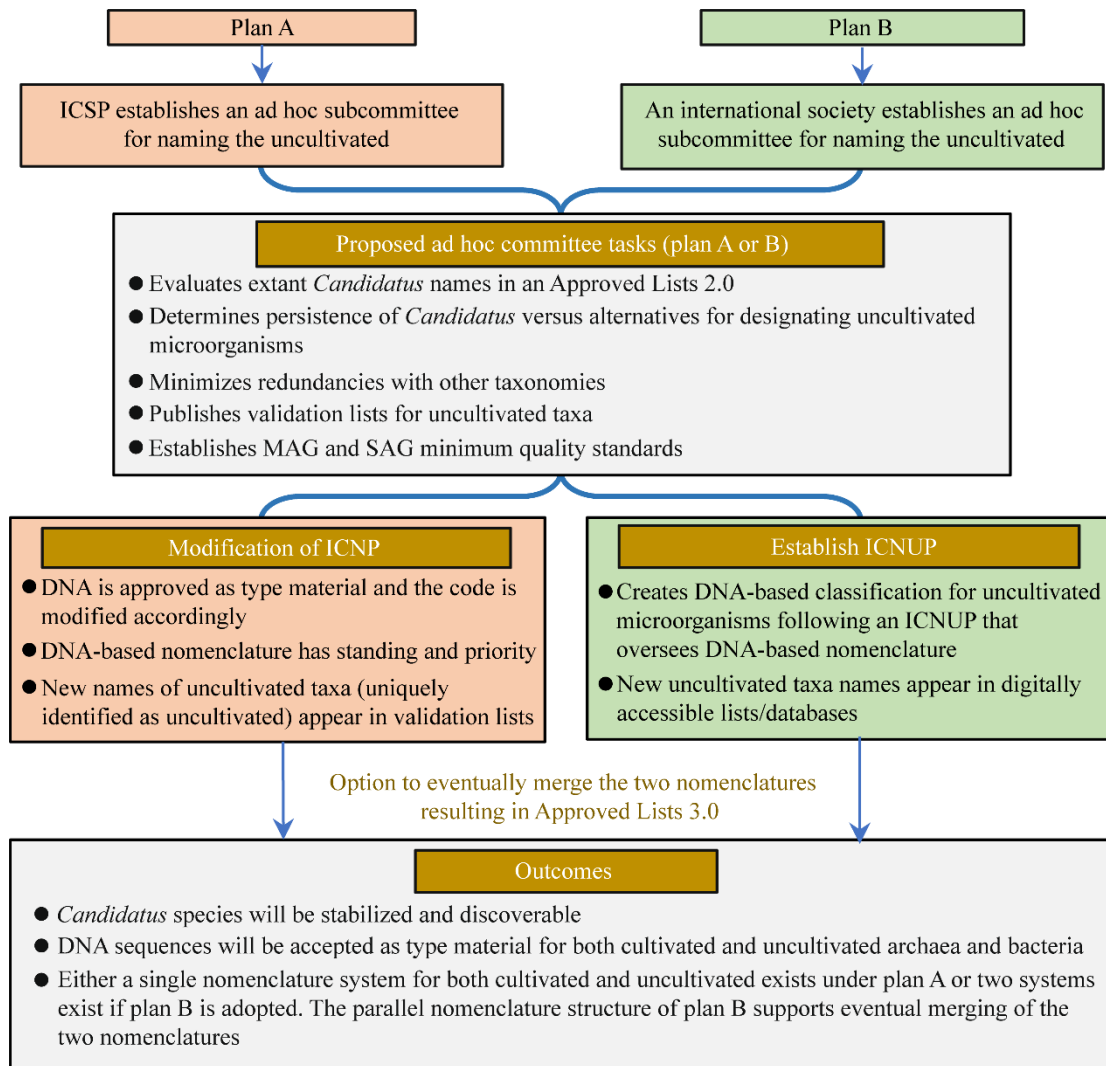


图 3 2020 年提出的未培养细菌与古菌命名系统的两套方案(引用并修改自文献[10])

Figure 3 Proposed roadmaps for nomenclature of uncultivated archaea and bacteria in 2020 (modified from reference [10]). Plans A and B provide two alternative paths for inclusion of uncultivated archaea and bacteria into the classical Linnaean nomenclature system. ICNP: International code of nomenclature of prokaryotes; ICNUP (latterly, the SeqCode): International code of nomenclature of uncultivated prokaryotes.

录并确立优先权。*Candidatus* 名称可以保留, 或者使用其他的标注方式来指示未培养状态。同时, 该组织可以在基因组标准协会(Genome Standards Consortium, GSC)的协助下, 制定相关数据的质量标准并提供针对未培养微生物的完整分类系统。ICNUP 的条款类似于方案 A 中的 ICNP, 已经发表的基于 DNA 序列信息的

Candidatus 名称将被确立优先权。方案 B 可能会与 ICNP 短期共存, 将来在学术界普遍支持的情况下, 两者将合并为一个统一的命名法规; 这两个命名法规也可能长期共存, 永远不会统一。方案 B 建议, 如果未培养类群被成功培养, 原先依据未培养法规创建的名称也应该被保留。

方案 A 希望通过修改 ICNP 来达到目标,可以避免不同系统在命名过程出现的矛盾、分歧和优先权问题,并且可以快速实施,能有效地满足当前学术界的需求。如果 ICSP 针对方案 A 作出决定所需要的时间过长,方案 B 就可以提供及时的解决方案,以避免原核生物命名系统中相关问题的加剧。然而,本质上这两种方案的终极目标是一致的,即希望未培养类群的名称唯一且稳定。

对于方案 A, ICNP 编委会在 2019 年秋季进行了多次讨论,随后 ICSP 执行委员会在 2020 年 1 月 2 日举行的会议上决定:先在 ICSP 会员中发起全面讨论(邮件形式),然后在 2020 年 3 月份对方案 A 的提议进行投票表决^[5]。此次投票除了决定是否将 DNA 序列纳入模式材料范围(含两项内容),也对另外两项提议进行表决:(1) 2020 年 1 月之前发表的符合 ICNP 规则的 *Candidatus* 名称是否享有优先权;(2) 是否使用类似于“^T” (标注在菌株号之后,表示模式/典型菌株)的上标“^{Ts}”或“^{Td}”标注未培养类群的 DNA 模式材料^[5]。共有 150 人参与投票前的讨论,最终确定了 23 人拥有表决权,结果于 2020 年 3 月 31 日进行统计。表决结果为以上 4 项提议均未通过。主要原因在于^[37]:(1) 对培养和保藏工作具有潜在的负面影响;(2) 被命名对象从生物体转向序列存在不确定风险;(3) 命名模式失去“可重复观测”的生物属性;(4) 可能会进一步增加基因组在分类学方法中的权重,不利于新方法的开发,并且与多相分类学(polyphasic taxonomy)的基本原则相背离;(5) 可能导致大量无意义、符号化名称的出现;(6) 导致模式的概念和类型模糊不清;(7) 无法有效解决“名古屋协定”等法规对生物多样性研究的限制;(8) 导致 ICNP 过多干预分类学自由。

于是,方案 B 正式启动。第一份命名法规

草案于 2020 年夏完成,被命名为 code of nomenclature of prokaryotes described from sequence data (SeqCode)。随后在 2021 年 2 月期间,由国际微生物生态学会(International Society for Microbial Ecology, ISME)组织了一系列线上研讨会,共有来自 42 个国家的 848 名会员参加,对 SeqCode 的内容进行讨论和修改(<https://www.isme-microbes.org/seqcode-workshops>)。第一版的 SeqCode 于 2022 年 9 月正式发布^[6]。

2.3 SeqCode 与 ICNP 的区别与联系

SeqCode 的出发点是将基因组序列认定为模式材料(也认定活的培养物作为模式),从而确保基于 MAG/SAG 发现的新类群和苛养菌能够被稳定、规范地命名^[7,10]。SeqCode 是基于 ICNP 的“演变”而非“变革”,更不是推倒重来。SeqCode 的主要特点包括^[9]:(1) 大部分条款与 ICNP 一致,为将来二者合并保留最大便利;(2) ICNP 有些条款之间存在矛盾,SeqCode 尝试化解矛盾;(3) 语句更直白、易懂,较少争议和混淆;(4) 通过线上注册和管理;(5) 为满足 ICNP 中 *Candidatus* 命名规则类群提供名称并赋予优先权。SeqCode 和 ICNP 的主要不同点如表 2 所示^[9]。单纯从生物命名的角度来讲,SeqCode 并未提出新的准则。例如,SeqCode 被争议最多的地方在于,不要求活培养物的生理学证据,而在 2001 年 1 月之前,ICNP 1990 年修订版对这一点也不做要求,仅是要求对命名模式进行详尽的描述和阐释即可^[8]。

2.3.1 模式

SeqCode 与 ICNP 最大的不同在于同时允许活的培养物和 DNA 序列(基因组序列)作为模式(表 2)^[7]。SeqCode 认为基因组序列同样可以满足作为模式的要求,可以提供分类单元存在的证据,并与其他分类单元的模式进行比较^[43]。此外,基因组序列作为模式比菌株有着多方面的

表 2 SeqCode 和 ICNP 的主要差异比较(引用并修改自文献[9])

Table 2 Comparison of the SeqCode and ICNP (modified from reference [9])

Category	ICNP	SeqCode	Comment
Type	Viable culture	DNA sequence and viable culture	The SeqCode recognizes ICNP names with priority
Criteria of authenticity	Viable culture representing type is deposited in two culture collections in different countries	Sequence representing type is deposited in an INSDC database and accession number is cited in effective publication and online registry	
Registration of names	Publication in <i>IJSEM</i> or Validation Lists	Online SeqCode Registry	SeqCode registration requires an effective publication
Compatibility of names with databases	None, requires manual curation	Database structure of registration system avoids a need of third parties and allows easy access, search and retrieval of information	
Governing body	International Committee on the Systematics of Prokaryotes (ICSP)	Committee on the Systematics of Prokaryotes Described from Sequence Data (CSPDSD)	
Number of taxonomic ranks above genus	11*	4	SeqCode does not recognize many rarely utilized taxonomic ranks
Type for taxonomic ranks above genus	Genus for all higher taxa [#]	Genus for all higher taxa	
Priority for taxonomic ranks above genus	Depends on date the name is validly published	Depends on the date the type genus is validly published	
Number of rules	65	50	Omission of rules related to unused taxa and names prior to the Approved List allowed reduction of SeqCode
Start date	1948 ^{&}	January 1, 2022	

*: Domain and kingdom are covered by ICNP after autumn of 2023^[14]; #: The type of class and subclass in ICNP was changed from order to genus in autumn of 2023^[14]; &: A draft code was prepared in 1936, and the first version of IBCN was finished in 1947 and formally published in 1948 (Figure 1).

优势^[7,17,44]: (1) 既适合未培养类群, 也适合可培养类群; (2) 储存便捷, 易于共享; (3) 可作为永久记录存放, 而菌株保藏需要定期活化且存活性难以保障。SeqCode 要求基因组数据达到最低质量标准^[6-7], 但该标准与 SeqCode 相互独立, 以适应不同时期的技术发展。因此, SeqCode 还需要一个组织来审查和制定序列数据的最低标准。对于一些特殊类群, 例如严格胞内寄生的支原体, 可制定专门的标准以保证可行性。

2.3.2 高级分类单元

由于 ICNP 的大部分条款是 20 世纪下半叶制定的, 那时候微生物的多样性是被严重低估的, 因此 ICNP 的内容并未很好地适应系统发育分类学。这导致在早期, 大量的属未被归类到更高一级的分类单元, 而纲则是作为最高级分类单元存在的。在 1980 年的认定名录中, 只有 7 个纲和 58 个科以上的名称^[9]。此外, ICNP 也没有制定纲的命名标准。在当时, ICNP 非常强调种和属的命名, 并且很少讨论更高分类等级

的命名规则。相比之下，目前科以上的分类单元约有 500 多个，纲有 100 多个^[45]。虽然 ICNP 2022 年修订版才正式纳入“门”^[13]，但是“门”在此之前就已经被广泛使用。由于早期 ICNP 未针对高级分类单元制定明确的规则，许多名称没有以统一的方式进行命名，从而导致了近些年高级分类单元名称的频繁变动，为学术界带来了诸多不便。

ICNP 中大部分高级分类单元(包括门)的模式是属^[13]，但是纲和亚纲的模式是目(已于 2023 年修改为属^[14])，这导致了整个分类系统名称的不连续性。ICNP 规定高级分类单元名称的优先权取决于名称的生效发表日期，却未规定与其模式对应名称的词源关联性^[46]，于是便会出现：即使其原来的模式属后来被判无效，一个有效发表的科及以上分类单元也拥有优先权^[47]。另外，ICNP 会导致有歧义的名称。例如，根据 ICNP 的规则，如果将 1980 年认定名录中的 *Alcaligenes* 和 1993 年命名的 *Burkholderia* 划归在同一个科，那么这个科的有效名称将是 1986 年命名的 *Alcaligenaceae*，而不是 2006 年命名的 *Burkholderiaceae*，因为 1986 年的 *Alcaligenaceae* 拥有优先权(发表时间更早)。然而，由于“*Alcaligenales*”从未被有效发表过，*Alcaligenaceae* 只能被归类为 *Burkholderiales*，于是从属到目便出现了名称的不连续性。

在 SeqCode 中，高级分类单元的优先权取决于所包含属的优先权^[7,9]。基本原则是，一个新属的建立同时意味着它可能代表一个新科、新目、新纲，甚至新门，即使这些高级分类单元当时未被命名。通过使用最早命名的属作为模式，以后不同属的合并就不会改变高级分类单元的名称，从而保证稳定性。如果高级分类单元被合并，则应该根据优先权更高(命名更早)的属名来选择或创建名称。如果一个高级分类

单元被分割，包含原来模式的分支将保留其名称。新建立的分支将根据所包含的属中拥有优先权(最早被命名)的属获得名称。一个全新的高级分类单元只能建立在同时发现一个新属的基础上。需要注意的是，SeqCode 接受 2022 年 1 月 1 日之前依据 ICNP 有效发布的所有名称^[7]。因此，SeqCode 中有关高级分类单元的命名规则，对该日期之前生效发表的高级分类单元无影响，只影响在该日期之后提出的高级分类单元名称。例如，根据现行 ICNP 和 SeqCode，任何包含 *Burkholderia* 属的目都将被命名为 *Burkholderiales*，这是因为 *Burkholderiales* 是在 2022 年之前根据 ICNP 规则命名的。假如是现在对包含 *Alcaligenes* 和 *Burkholderia* 的目进行命名，根据 SeqCode 的规则，“*Alcaligenales*”才是正确的名称，因为 *Alcaligenes* 是更早被命名的属。

2.3.3 名称类别

ICNP 将名称分为 5 类：合法(legitimate)、非法(illegitimate)、有效发表(effectively published)、生效发表(validly published)和正确(correct)(图 4A)。有效发表是指在正规的出版物上发表，包含其他 4 类名称。合法名称是根据 ICNP 规则形成的，必须是生效发表。生效发表是指在 *IJSEM* 或其认定名录中发表。虽然原则上生效发表的名称都应该是合法的，但也有例外。例如有些生效发表的名称会因为分类单元合并时违反了优先权原则，或者因为命名法规的变更而变得不合法。最后，正确名称必须是合法的，但是只针对分类时使用，即现行分类系统中有效的、被认可的名称。例如，一个名称本来是合法的，但是可能会因为再分类而在分类系统中不再是正确的。这些区别在 SeqCode 中也有使用。然而，与 ICNP 不同的是，SeqCode 使用在线注册系统，即 SeqCode Registry (<https://disc-genomics.uibk.ac.at/seqcode/>)^[6-7]。在这个系统

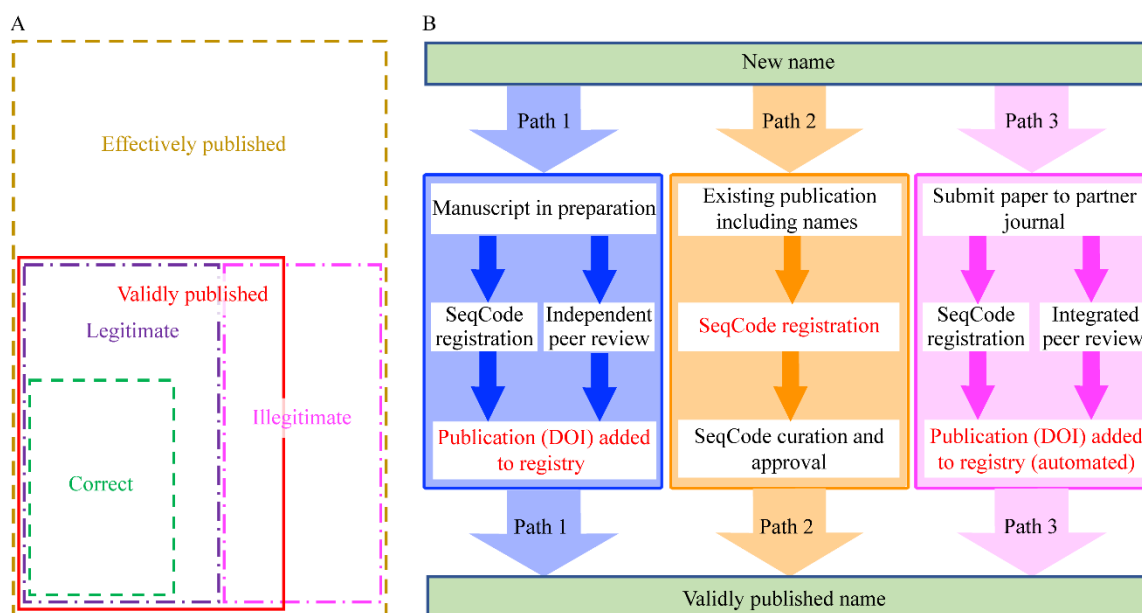


图 4 名称的类型(A)和 SeqCode 中名称的生效发表过程(B) (引用并修改自文献[6])

Figure 4 Categories of names (A) and validation process of names under the SeqCode (B) (modified from reference [6]). Path 3 has not been implemented in the SeqCode registry. Steps marked in red indicate where the date of priority was marked.

中, 由作者完成新名称的注册, 最好是在论文发表之前完成。这些名称将由管理员进行审查和批准。名称在有效发表和完成相应模式序列提交之后, 将成为生效发表。SeqCode 通过该系统中名称的生效发表日期确定优先权, 而不是依据期刊上的发表日期, 从而保证在任何期刊上发表名称享有均等的机会。此外, SeqCode Registry 将为尚未生效发表的名称保留一年时间, 以避免后续出版物中出现同义名称。

2.3.4 生效发表要求

SeqCode Registry 的目标是提供一个与元数据相关联的名称列表, 这些名称需要满足 SeqCode 的模式数据标准和正字法。该系统不会制定任何分类原则, 但是仍然会针对如何为未培养类群命名提供建议^[48]。同样, ICNP 对名称的认定也只有很少的要求, 即: 符合正字法; 有效发表; 来自两个国家的两份保藏证明。因此, 即使是使用计算机软件创建数以千计的名称, 同时相

关生物类群的生物属性几乎未被研究^[23], SeqCode 也不会干预这些论文的发表。然而, 在缺乏生物学分析的情况下, 应该尽量避免这样的命名。

2.3.5 分类等级

在历史上, 各分类等级的亚单元曾广泛被使用, 例如亚纲、亚目等, 各版本 ICNP 使用的分类等级多达 15 个: 亚种(subspecies)、种(species)、亚属(subgenus)、属(genus)、亚族(subtribe)、族(tribe)、亚科(subfamily)、科(family)、亚目(suborder)、目(order)、亚纲(subclass)、纲(class)、门(phylum)、界(kingdom)和域(domain)^[8,13-14]。由于概念模糊和实验证据不足, 目前亚单元已很少被讨论和使用, 但是 ICNP 2022 年修订版仍然包含 13 个分类等级(不含亚族和亚科)^[14]。而 SeqCode 仅包含 7 个分类等级: 亚种、种、属、科、目、纲和门^[9]。

2.3.6 生效命名途径

基于 SeqCode 的名称可以在 SeqCode Registry

系统 (<https://disc-genomics.uibk.ac.at/seqcode/>) 进行生效验证^[7]。在该系统中可进行名称和模式的注册、验证和原始数据查询。SeqCode Registry 有 3 个主要目标：(1) 对根据 SeqCode 创建的名称进行注册和评估；(2) 对文献中使用的名称进行自动识别，以保证这些名称的规范化和标准化；(3) 维护依据 SeqCode 生效发表名称的合法性，发布相应的名录，提供模式序列的原始数据链接。目前，可以使用两种方式注册和验证名称(图 4B)，第三种方式有望在将来实现^[6-7]。推荐的方式是：在相关论文投稿前，也就是论文发表前就进行名称的注册和审查。SeqCode Registry 将自动检查并管理提交的名称。

途径 1 主要处理以下两种情况。(1) 在预注册期间进行自动检查以防止在名称发表之前出现同名或语法错误等问题，从而减少发布后的名称更改现象。该过程类似于 *IJSEM* 同行评审期间人工进行的名称检查。SeqCode Registry 将尽可能通过自动化流程，加快处理速度并减少人为错误。同样，数据质量检查是为了保证作为命名模式的基因组数据的质量。(2) 在预注册过程中生成的 SeqCode ID，也就是统一资源定位符(uniform resource locators, URL)，可以标注在提交的稿件中促成有效发表。这些 URL 对审稿人和编辑开放访问，以验证相关名称是否通过了 SeqCode 检查。这一过程将为新名称和相关基因组的同行评审提供便利，因为 SeqCode Registry 在预注册阶段就可以保证这些名称没有问题且作为模式的序列数据质量达标。在途径 1 中，用户提供论文的数字对象标识符(digital object identifier, DOI)之后，注册过程才算完成，这也是名称的优先权日期。这一步通常由作者完成，如果在有效发表的论文中使用了 URL，SeqCode Registry 也将自动捕获 DOI，从而完成注册过程。

途径 2 用于注册和验证已经发表的名称，包括 *Candidatus* 名称。将名称和原始数据输入到 SeqCode Registry 中后，系统将进行与途径 1 相同的自动检查。然后，管理员对提交的名称和论文接收情况进行审查，从而完成注册，并标记名称的优先权日期。由于名称或基因组数据质量在论文发表之前未经过审核和纠正，途径 2 通常不如途径 1 理想，因为根据 SeqCode 的规定，已发表的 *Candidatus* 名称可能会被判定无效(如需要修正或无优先权等)。途径 3 致力于推动论文的同行评审和名称的注册、审核流程同时进行，类似于 *IJSEM* 对基于 ICNP 名称的综合评审系统，但是这需要相关期刊的积极配合并重组编辑委员会。

2.4 SeqCode 的争议之处

毫无疑问，SeqCode 将基因组序列作为命名模式之后，新发现的未培养微生物便可以获得规范、生效的拉丁名称，这对原核生物多样性的研究和交流活动具有积极促进作用。然而，ICSP 通过投票拒绝了将基因组序列作为模式纳入 ICNP 中的提议。本文将持反对意见的专家所关心的问题总结为以下 5 点^[5]。

2.4.1 重复命名

从当前科学研究成果发表和传播的模式来看，重复命名现象不可避免。例如，不同学者在同一段时间内发表的名称，由于论文审稿、发表和名称收录、检索等环节的时效性问题，彼此之间很难直接进行比较和验证，同义名称就可能出现。在 ICNP 规定范围内，每年新发表的名称数量有限(1 000 个左右)，重复命名概率较低，也易于修正。当 SeqCode 实施后，可以预见短期内会出现大量的新名称。因为发表时效性问题，未培养和可培养类群名称的重复性会增加。另一方面，细菌与古菌基因组百科全书(genomic encyclopedia of bacteria and archaea,

GEBA)计划的实施^[49], 让原核生物分类学正式进入基因组时代, 用基因组学方法构建更为理想的分类体系成为现实。然而, 仍然有相当比例的模式菌株未进行基因组测序, 这将使部分通过 MAG/SAG 发现的类群难以界定是否属于新类群, 这也增加了重复命名的可能性。

虽然 SeqCode 提出了处理自身与 ICNP 名称优先权问题的解决方案, 但是 ICNP 和 *IJSEM* 对未培养类群名称的处理方式并没有改变^[50]。例如, 根据双方的规定: 依据 SeqCode 在 ICNP 的属内命名新种是可行的(不会引起重复命名和误解); 但是, 依据 ICNP 在 SeqCode 的属内命名新种却无法实现生效发表, 因为 ICNP 不认可基于 SeqCode 命名的属名; 基于 SeqCode 已被命名的类群, 后期获得了纯培养, 即使不使用基于 SeqCode 的名称, 也能实现基于 ICNP 的生效发表。对于属以上的高级分类单元, 情况更为复杂。在 SeqCode 的名称系统中, 一个或多个属被命名后, 根据优先权规则, 与它们相关的高级分类单元的名称便是确定的, 无论在当时有没有被命名。然而, 在 ICNP 中高级分类单元的名称需要被正式命名才能确定, 并且不要求与其中命名最早的属的名称在词源上有关联。另外, 由于 ICNP 不认可 SeqCode 中的属名, 这会导致同一个高级分类单元的模式属在两个命名系统中不同, 进一步导致所有相关高级分类单元的名称也不相同。针对这些问题, SeqCode 尚未提出很好的解决方案。

2.4.2 限制微生物资源发掘

在经典分类学中, 能够被纯培养的菌种是进行研究的基本材料。ICNP 要求纯培养物来进行命名有助于鼓励分离纯培养物, 从而促进菌种在农业、医药等领域的利用, 例如疫苗开发、各领域的菌剂研制等。SeqCode 将基因组序列作为命名模式, 那么可能导致的一种结果是: 即使是

获得了纯培养物, 部分学者也不会将其用作模式。因为将菌株存放到保藏中心比将基因组序列上传至公共数据库的流程要烦琐得多, 而上传序列对 ICNP 和 SeqCode 都是必须完成的。另外, 一旦发现获得纯培养所代表的类群已经依据 SeqCode 被命名, 研究者进行后续研究并将其发表的热情也会大打折扣。也有学者认为, 获得纯培养后相关分类单元的描述便有必要被修订, 因而培养和保藏的动机仍然存在。然而修订与重新命名的学术价值权重是不同的, 事实上专门修订分类单元的研究在实践中确实相对较少。虽然在 SeqCode 的影响下, 分离菌种在某些领域的研究中仍然存在明确的需求, 但是命名法规本身应该鼓励菌种的分离和培养。

2.4.3 模式“可重复观测”的属性消失

命名模式的一个基本属性是可以被反复观测、研究, 并且随着技术方法的进步可以通过对模式的再次研究得到新发现, 从而不断加深对生物类群的认识^[9]。菌株、标本、化石和稳定的共培养物和 DNA 分子均具有这一属性。然而, 基因组序列却不具备这一属性。严格来讲, 作为模式的基因组序列与菌株完全不同, 它不是一种生理学材料, 只是在特定时期通过特定手段获得的一种实验结果^[5]。虽然 SeqCode 要求公开原始数据(raw reads), 但是这些数据在相应的菌株被培养之前将无法更新, 因技术原因丢失的生物学信息将无法补救, 即无法通过“重复观测”获得新认知。另一方面, 单独通过基因组序列对生物体进行认知很可能会得到片面的结论, 许多推测得到的表型特征和生态功能仍需要使用培养物进行验证。从这方面看, SeqCode 对整个微生物学领域的研究可能会产生负面影响。

2.4.4 自动化命名与命名原则的背离

合理、恰当的名称可以反映生物类群的典型特征, 在记载和交流时易于理解、方便记忆,

这是对生物体进行命名时的根本出发点和首要原则^[13]。毫无疑问,将基因组序列作为命名模式后,一些研究将发现大量需要被命名的新分类单元。例如,2023年的两份研究分别一次性报道了65 000多个和329个名称^[37,51]。由于缺少系统而全面的认识,为所有新类群创建有意义、合理的名称将变得极为困难。为了解决批量命名过程中的语法问题,一款Python程序GAN (great automatic nomenclator)应运而生^[20]。然而,这也仅能保证这些名称没有语法错误,如果不能反映被命名生物体的生物学特性,那么相比于“字母+数字”的形式和*Candidatus*名称,这种名称并没有多大优势。另外,自动化命名也很难保证不会与生物体的特征相冲突,如果相关类群后来成功被培养,却只能使用具有误导性的名称,这无疑也违背了命名的初衷。一般来说,只有在对分类单元有所了解的时候才需要进行命名。当然,如果考虑到命名或提供标签主要是为了在开展相关研究时方便交流与记载,那么计算机生成的、缺乏指示意义的名称也具有价值。无论如何,命名都有助于追溯相关类群已获得的知识,但是命名应该是对生物体进行研究的起点而不是终点^[9]。

2.4.5 推广和应用体系尚不成熟

SeqCode的定位是ICNP的补充和扩展,在两者并行发展过程中,SeqCode力求不与ICNP产生较大矛盾^[6,9]。例如,一旦相关类群被成功培养并满足ICNP对其模式的要求,SeqCode期望相关名称能顺利被ICNP接受。然而,这一目标的实现,尚有一些实际问题需要解决。SeqCode主要是由分子生态学家提出和推广,在缺少学术出版物和相关组织明确引导的情况下,传统分类学家对SeqCode的了解和接纳程度如何,尚有待验证。另一方面,SeqCode尚未提供简单快捷的方法或工具来帮助分类学家判定新

分类单元。虽然这并不是命名法规的职责,但是SeqCode和ICNP的双重影响势必导致新分类单元的识别难度增加。这对分类学家提出了更高的要求,增加了名称系统的不确定性和未来SeqCode与ICNP在融合过程中的困难^[7]。

3 结论与展望

近一个世纪以来,ICNP不断改进、完善,很好地促进了原核生物分类学的发展。在过去的5年里,ICSP接连发布了ICNP 2008年和2022年两个修订版^[8,13],也针对部分条款进行了多次小范围的订正(图1)。这表明,ICNP在通过不断改进以适应组学时代原核生物命名的需求。另一方面,伴随着基因组测序技术的快速发展和海量未培养新类群的发现,基于基因组序列为新类群命名的需求日益增加。2020年,ICSP通过投票表决正式拒绝了将基因组序列作为模式纳入ICNP的提议,这直接导致了SeqCode的出现。SeqCode致力于解决未培养类群命名的规范性和与ICNP的兼容性问题。虽然SeqCode尽量避免和ICNP产生较大分歧,并保留将来与ICNP合并的可能性,但是两套命名法规依然存在一定的竞争关系^[7]。目前,*IJSEM*对基于基因组序列的名称,不论是否基于SeqCode命名,仍然按照ICNP中关于*Candidatus*的规定来处理^[50],也未对相关描述和模式数据的质量要求进行说明。另一本知名分类学期刊*Systematic and Applied Microbiology*则针对基于SeqCode的未培养类群命名和描述所需要满足的条件进行了明确规定^[52],其他学术期刊尚未发布相关公告。由此可见,SeqCode对未培养类群命名的作用效力仍有待观察。

作为原核生物命名法的根基,ICNP在面对未知风险时采取保守的措施,拒绝将基因组序列作为模式纳入其中,可以尽可能确保现有的名

称系统和分类学体系维持稳定。作为 ICNP 在未培养类群命名方面的补充, SeqCode 无疑有利于实现基于基因组序列进行命名的标准化和规范化, 能够促进人们对微生物多样性的认识。可以预见在 SeqCode 和 ICNP 共存期间, 最为直接和明显的问题便是“重复命名”和 SeqCode 的“推广和应用体系不成熟”。两套命名法规也势必导致分类学家基于纯培养(ICNP 命名体系)和未培养数据库(SeqCode 命名体系)鉴定新类群的工作量加大、难度增加。另一方面, 有多少传统分类学家会关注和遵守 SeqCode, 目前仍难以确定^[7]。这也会增加重复命名的风险, 给整个原核生物名称系统的稳定性带来了不确定因素。因此, 作者呼吁相关微生物学者共同关注 ICNP 和 SeqCode 命名体系(LPSN, *IJSEM* 和 SeqCode Registry 等)及相关动态, 本着严谨的治学态度, 务实的科学精神开展微生物多样性研究工作, 以期通过共同努力构建更加合理、有效的原核生物名称系统, 服务广大学术界。

致谢

感谢美国佐治亚大学微生物学系的 William B. Whitman 教授对本文图表和英文摘要提出的宝贵意见。

参考文献

- [1] ICSP. International Committee on Systematics of Prokaryotes[Z]. [2023-09-20]. <https://www.the-icsp.org/index.php>.
- [2] JIAO JY, LIU L, HUA ZS, FANG BZ, ZHOU EM, SALAM N, HEDLUND BP, LI WJ. Microbial dark matter coming to light: challenges and opportunities[J]. *National Science Review*, 2021, 8(3): nwaa280.
- [3] MEZITI A, RODRIGUEZ-R LM, HATT JK, PEÑAGONZALEZ A, LEVY K, KONSTANTINIDIS KT. The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample[J]. *Applied and Environmental Microbiology*, 2021, 87(6): e02593-20.
- [4] MOSS EL, MAGHINI DG, BHATT AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing[J]. *Nature Biotechnology*, 2020, 38: 701-707.
- [5] SUTCLIFFE IC, DIJKSHOORN L, WHITMAN WB, EXECUTIVE BOARD OBOTI. Minutes of the International Committee on Systematics of Prokaryotes online discussion on the proposed use of gene sequences as type for naming of prokaryotes, and outcome of vote[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2020, 70(7): 4416-4417.
- [6] HEDLUND BP, CHUVOCHINA M, HUGENHOLTZ P, KONSTANTINIDIS KT, MURRAY AE, PALMER M, PARKS DH, PROBST AJ, REYSENBACH AL, RODRIGUEZ-R LM, ROSSELLO-MORA R, SUTCLIFFE IC, VENTER SN, WHITMAN WB. SeqCode: a nomenclatural code for prokaryotes described from sequence data[J]. *Nature Microbiology*, 2022, 7: 1702-1708.
- [7] WHITMAN WB, CHUVOCHINA M, HEDLUND BP, KONSTANTINIDIS KT, PALMER M, RODRIGUEZ-R LM, SUTCLIFFE I, WANG F. Why and how to use the SeqCode[J]. *mLife*, 2024, 3: 1-13.
- [8] PARKER CT, TINDALL BJ, GARRITY GM. International Code of Nomenclature of Prokaryotes prokaryotic code (2008 Revision)[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2019, 69(1a): S7-S111.
- [9] WHITMAN WB, CHUVOCHINA M, HEDLUND BP, HUGENHOLTZ P, KONSTANTINIDIS KT, MURRAY AE, PALMER M, PARKS DH, PROBST AJ, REYSENBACH AL, RODRIGUEZ-R LM, ROSSELLO-MORA R, SUTCLIFFE I, VENTER SN. Development of the SeqCode: a proposed nomenclatural code for uncultivated prokaryotes with DNA sequences as type[J]. *Systematic and Applied Microbiology*, 2022, 45(5): 126305.
- [10] MURRAY AE, FREUDENSTEIN J, GRIBALDO S, HATZENPICHLER R, HUGENHOLTZ P, KÄMPFER P, KONSTANTINIDIS KT, LANE CE, PAPKE RT, PARKS DH, ROSSELLO-MORA R, STOTT MB, SUTCLIFFE IC, THRASH JC, VENTER SN, WHITMAN WB, ACINAS SG, AMANN RI, ANANTHARAMAN K, ARMENGAUD J, et al. Roadmap for naming uncultivated archaea and bacteria[J]. *Nature Microbiology*, 2020, 5: 987-994.

- [11] OREN A, ARAHAL DR, GÖKER M, MOORE ERB, ROSSELLO-MORA R, SUTCLIFFE IC. Proposals to emend Rules 8, 15, 22, 25a, 30(3) (b), 30(4), 34a, and Appendix 7 of the International Code of Nomenclature of Prokaryotes[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2022, 72(12): 005630.
- [12] OREN A, GARRITY GM. Valid publication of the names of forty-two phyla of prokaryotes[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2021, 71(10): 005056.
- [13] OREN A, ARAHAL DR, GÖKER M, MOORE ERB, ROSSELLO-MORA R, SUTCLIFFE IC. International Code of Nomenclature of Prokaryotes. Prokaryotic code (2022 revision)[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2023, 73(5a): 005585.
- [14] OREN A. Emendation of Principle 8, Rules 5b, 8, 15, 33a, and Appendix 7 of the International Code of Nomenclature of Prokaryotes to include the categories of kingdom and domain[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2023, 73(11): 006123.
- [15] WINSTON JE. Twenty-first century biological nomenclature: the enduring power of names[J]. *Integrative and Comparative Biology*, 2018, 58(6): 1122-1131.
- [16] LAPAGE SP, SNEATH PHA, LESSEL EF, SKERMAN VBD, SEELIGER HPR, CLARK WA. International Code of Nomenclature of Bacteria[M]. Washington, D.C.: American Society for Microbiology, 1992.
- [17] PALMER M, SUTCLIFFE I, VENTER SN, HEDLUND BP. It is time for a new type of type to facilitate naming the microbial world[J]. *New Microbes and New Infections*, 2022, 47: 100991.
- [18] PALLEN MJ. Bacterial nomenclature in the era of genomics[J]. *New Microbes and New Infections*, 2021, 44: 100942.
- [19] TRÜPER H. How to name a prokaryote? Etymological considerations, proposals and practical advice in prokaryote nomenclature[J]. *FEMS Microbiology Reviews*, 1999, 23(2): 231-249.
- [20] PALLEN MJ, TELATIN A, OREN A. The next million names for archaea and bacteria[J]. *Trends in Microbiology*, 2021, 29(4): 289-298.
- [21] OREN A. Prokaryotic nomenclature[M]//Bergey's Manual Trust. *Bergey's Manual of Systematics of Archaea and Bacteria*: Wiley Online Library, 2019: 1-12.
- [22] OREN A. How to name new taxa of archaea and bacteria[M]//Bergey's Manual Trust. *Bergey's Manual of Systematics of Archaea and Bacteria*: Wiley Online Library, 2019: 1-24.
- [23] OREN A. Naming new taxa of prokaryotes in the 21st century[J]. *Canadian Journal of Microbiology*, 2023, 69(4): 151-157.
- [24] OREN A, SCHINK B, GARRITY GM. Wanted: microbiologists with basic knowledge of Latin and Greek to join our 'nomenclature quality control' team[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2015, 65(11): 3761-3762.
- [25] ROSSELLÓ-MÓRA R, WHITMAN WB. Dialogue on the nomenclature and classification of prokaryotes[J]. *Systematic and Applied Microbiology*, 2019, 42(1): 5-14.
- [26] FOX GE, STACKEBRANDT E, HESPELL RB, GIBSON J, MANILOFF J, DYER TA, WOLFE RS, BALCH WE, TANNER RS, MAGRUM LJ, ZABLEN LB, BLAKEMORE R, GUPTA R, BONEN L, LEWIS BJ, STAHL DA, LUEHRSEN KR, CHEN KN, WOESE CR. The phylogeny of prokaryotes[J]. *Science*, 1980, 209(4455): 457-463.
- [27] MURRAY RG, SCHLEIFER KH. Taxonomic notes: a proposal for recording the properties of putative taxa of prokaryotes[J]. *International Journal of Systematic Bacteriology*, 1994, 44(1): 174-176.
- [28] MURRAY RG, STACKEBRANDT E. Taxonomic note: Implementation of the provisional status *Candidatus* for incompletely described prokaryotes[J]. *International Journal of Systematic Bacteriology*, 1995, 45(1): 186-187.
- [29] OREN A, GARRITY GM, TRUJILLO ME. Registration of names of prokaryotic *Candidatus* taxa in the *IJSEM*[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2020, 70(7): 3955.
- [30] OREN A, GARRITY GM, PARKER CT, CHUVOCHINA M, TRUJILLO ME. Lists of names of prokaryotic *Candidatus* taxa[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2020, 70(7): 3956-4042.
- [31] OREN A. *Candidatus* List No. 4: Lists of names of prokaryotic *Candidatus* taxa[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2022, 72(11): 005545.
- [32] OREN A, GARRITY GM. *Candidatus* List No. 3. Lists of names of prokaryotic *Candidatus* taxa[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2022, 72(1): 005186.
- [33] OREN A. Nomenclature of prokaryotic '*Candidatus*' taxa: establishing order in the current chaos[J]. *New Microbes and New Infections*, 2021, 44: 100932.

- [34] OREN A. A plea for linguistic accuracy-also for *Candidatus* taxa[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2017, 67(4): 1085-1094.
- [35] OREN A, GARRITY GM. *Candidatus* List No. 2. Lists of names of prokaryotic *Candidatus* taxa[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2021, 71(4): 004671.
- [36] PARKS DH, CHUVOCHINA M, RINKE C, MUSSIG AJ, CHAUMEIL PA, HUGENHOLTZ P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy[J]. *Nucleic Acids Research*, 2022, 50(D1): D785-D794.
- [37] PALLEN MJ, RODRIGUEZ-R LM, ALIKHAN NF. Naming the unnamed: over 65 000 *Candidatus* names for unnamed archaea and bacteria in the genome taxonomy database[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2022, 72(9): 005482.
- [38] HEDLUND BP, DODSWORTH JA, STALEY JT. The changing landscape of microbial biodiversity exploration and its implications for systematics[J]. *Systematic and Applied Microbiology*, 2015, 38(4): 231-236.
- [39] KONSTANTINIDIS KT, ROSSELLÓ-MÓRA R. Classifying the uncultivated microbial majority: a place for metagenomic data in the *Candidatus* proposal[J]. *Systematic and Applied Microbiology*, 2015, 38(4): 223-230.
- [40] WHITMAN WB. Modest proposals to expand the type material for naming of prokaryotes[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2016, 66(5): 2108-2112.
- [41] WHITMAN WB, SUTCLIFFE IC, ROSSELLO-MORA R. Proposal for changes in the international code of nomenclature of prokaryotes: granting priority to *Candidatus* names[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2019, 69(7): 2174-2175.
- [42] KONSTANTINIDIS KT, ROSSELLÓ-MÓRA R, AMANN R. Uncultivated microbes in need of their own taxonomy[J]. *The ISME Journal*, 2017, 11(11): 2399-2406.
- [43] WHITMAN WB. Genome sequences as the type material for taxonomic descriptions of prokaryotes[J]. *Systematic and Applied Microbiology*, 2015, 38(4): 217-222.
- [44] KONSTANTINIDIS KT, ROSSELLÓ-MÓRA R, AMANN R. Advantages outweigh concerns about using genome sequence as type material for prokaryotic taxonomy[J]. *Environmental Microbiology*, 2020, 22(3): 819-822.
- [45] PARTE AC, SARDÀ CARBASSE J, MEIER-KOLTHOFF JP, REIMER LC, GÖKER M. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2020, 70(11): 5607-5612.
- [46] LUDWIG W, EUZÉBY J, WHITMAN WB. Taxonomic outline of the phylum *Actinobacteria*[M]//*Bergey's Manual of Systematic Bacteriology*. New York: Springer, 2012: 29-31.
- [47] WHITMAN WB, LAWSON PA, LOSEY NA. Response to Tindall (2014) on the legitimacy of the names *Solimonadaceae* Losey et al. 2013, *Xanthomonadaceae* Saddler and Bradbury 2005 and *Xanthomonadales* Saddler and Bradbury 2005[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2015, 65(3): 1086-1087.
- [48] CHUVOCHINA M, RINKE C, PARKS DH, RAPPÉ MS, TYSON GW, YILMAZ P, WHITMAN WB, HUGENHOLTZ P. The importance of designating type material for uncultured taxa[J]. *Systematic and Applied Microbiology*, 2019, 42(1): 15-21.
- [49] WU DY, HUGENHOLTZ P, MAVROMATIS K, PUKALL R, DALIN E, IVANOVA NN, KUNIN V, GOODWIN L, WU M, TINDALL BJ, HOOPER SD, PATI A, LYKIDIS A, SPRING S, ANDERSON IJ, D'HAESELEER P, ZEMLA A, SINGER M, LAPIDUS A, NOLAN M, et al. A phylogeny-driven genomic encyclopaedia of bacteria and archaea[J]. *Nature*, 2009, 462: 1056-1060.
- [50] GÖKER M, MOORE ERB, OREN A, TRUJILLO ME. Status of the SeqCode in the *International Journal of Systematic and Evolutionary Microbiology*[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2022, 72(12): 005754.
- [51] CHUVOCHINA M, MUSSIG AJ, CHAUMEIL PA, SKARSHEWSKI A, RINKE C, PARKS DH, HUGENHOLTZ P. Proposal of names for 329 higher rank taxa defined in the Genome Taxonomy Database under two prokaryotic codes[J]. *FEMS Microbiology Letters*, 2023, 370: fnad071.
- [52] ROSSELLÓ-MÓRA R, KONSTANTINIDIS KT, AMANN R. How *Systematic and Applied Microbiology* will deal with two nomenclature codes (ICNP and SeqCode) for prokaryotes, and which classification standards are recommended for new taxa descriptions[J]. *Systematic and Applied Microbiology*, 2023, 46(1): 126371.