

规范基因工程文稿常见符号的编排

贺 密 青

中国科技出版传媒股份有限公司,100717,北京

摘 要 针对基因工程中涉及的基因及其供体与受体、工具酶、载体等符号的大小写和正斜体编排进行分析,认为作为基因的供体与受体:病毒各级分类单位均应编排成斜体,暂定种编排成正体,种名中的专有名词大写;亚属名编排成斜体,首字母大写,等级指定词用正体;品种加词放在单引号内,每个词都用正体,首字母大写。基因符号的所有组成部分均编排成斜体,根据基因的物种来源判断基因符号的大小写。T4、T7等DNA连接酶,T4、T7编排成正体,4、7平排;Taq DNA聚合酶中的“Taq”编排成斜体。质粒载体上的基因符号编排成小写;质粒的命名可加上相应的缺失基因或外源基因,将基因符号编排成斜体。以上结论希望能为规范基因工程文稿的编排提供参考。

关键词 基因工程;符号编排;大小写;正斜体

Editorial standardization of common symbols in genetic engineering papers//HE Yaoqing

Abstract In this paper, the capital and small case, upright and italic arrangement of the symbols of gene donor and receptor, gene, instrumental enzyme and vector involved in genetic engineering are analyzed. In order to standardize the symbol layout of genetic engineering papers, we propose that: 1) as the donor and receptor of gene, the taxonomic units of virus at all levels should be arranged in italics, the tentative species should be arranged in upright, and the proper nouns in the species name should be capitalized; 2) the subgenus names are arranged in italics, the first letter is capitalized, and the words of grade specifier are in upright; 3) varieties plus words are put in single quotation marks, and each word is capitalized; 4) all the components of the gene symbol are arranged in italics, and the capital and small case of the gene symbol is judged according to the species source of the gene; 5) DNA ligases such as T4 and T7, T4 and T7 were arranged into upright, and 4 and 7 were arranged flat row; 6) the “Taq” in Taq DNA polymerase is arranged in italics; 7) the gene symbols on plasmid vector are arranged in lowercase; 8) the name of plasmid can be added with the corresponding deletion gene or foreign gene, and the gene symbol can be arranged in italics.

Keywords genetic engineering; symbol layout; capital and small case; upright and italic

Author's address China Science Publishing & Media Ltd., 100717, Beijing, China

DOI:10.16811/j.cnki.1001-4314.2021.04.009

在微生物遗传学和分子生物学基础上形成的基因工程,是在工具酶的作用下,将供体生物的目的基因与

作为载体的DNA分子连接起来,然后与载体一起导入受体生物中,让目的基因在其中正常复制和表达的一种崭新技术。这种崭新技术不但可以用来培养高产、优质、抗性好的动植物新品种,而且可以用来研究致病基因的致病机制、开发基因工程药物和疫苗、创建基因诊断和治疗技术等。近些年,有关基因工程的稿件逐年增加,但笔者在编辑此类稿件时经常可见与基因工程有关的常见符号的大小写和正斜体编排不规范的问题,这不仅给读者带来阅读障碍,而且影响学科之间的交流和研究成果的推广,有时甚至引发误解。例如,在生物学领域,蛋白质符号用正体编排,与其对应的基因符号用斜体编排,如果两者混淆,就会影响科学概念的表达,如在“表明MeCP2可能是一种转录因子”这句话中,将“MeCP2”排成了斜体,让人误解“MeCP2”是基因,其实转录因子是指能识别启动子、增强子或特定序列而调控基因表达的蛋白质^[1],不是基因,应该编排成正体。鉴于学术语言不仅仅指纯文字的语言表达,在自然科学中,经常使用的还有符号语言^[2]。同时,在基因工程这项技术中,基因的供体与受体、基因、工具酶、载体等符号的编排都有一定的规则。因此,本文对稿件中常见基因工程符号的编排进行分析,并提出建议,以期为规范基因工程文稿的编排提供参考。

1 基因的供体与受体

能够提供功能基因的生物体就是基因的供体,能够扩增或表达功能基因的生物体就是基因的受体。随着基因工程的发展,自然界的生物,即微生物、植物、动物等物种的细胞都可以作为基因工程的供体或受体细胞,这就需要规范以拉丁文符号表示的物种名称的编排。《国际细菌命名法规》《国际植物命名法规》《国际栽培植物命名法规》《国际动物命名法规》对此进行了规范,主要为属以上分类单位用正体表示,属和种学名用斜体表示,但属以上学名首字母大写,种学名首字母小写,命名人用正体,首字母大写。例如,规范编排为:禾本科Gramineae,玉蜀黍属Zea,玉蜀黍Zea mays Linn.;不应编排成Gramineae、Zea,以及Zea Mays Linn.、Zea mays Linn.、zea mays Linn.等。相关文献依据瑞典植物学家林奈的“双名制”对此也进行了探讨^[3-6],主要强调以下几点。

1.1 病毒

病毒的命名规则与其他物种不同,其种名一般为英文单词,在一般性叙述中不使用斜体,仅在指代分类系统中的正式种名时使用斜体^[4-5],但时常可见如疱疹病毒科(Herpesviridae),伪狂犬病毒(Porcine Pseudorbies Virus)^[7],或 Porcine pseudorbies virus(编排成正体,会被误认为是暂定名)等编排方式,均不规范。

《国际病毒分类与命名原则》规定,病毒分类系统的目、科、亚科、属名均编排成斜体,且首字母大写,如套式病毒目(Nidovirales),冠状病毒科(Coronaviridae),冠状病毒属(Coronavirus)。在种名的编排上,1998年国际病毒分类委员会在第27次常务会上提出所有的病毒种名用斜体,第一个词的首字母大写,其他词除专有名词外首字母一般不大写^[8-10]。例如,啤酒酵母20S裸露RNA病毒(Saccharomyces 20S RNA narnavirus)的所有字母均为斜体,包括数字20和单字母S;RNA是专有名词,编排为大写、斜体。此次会议同时还规定,暂定种不用斜体,首字母大写^[8-10],如ΦKMV噬菌体属(PhiKMV-like viruses)^[11]。

1.2 亚分类单位

存在亚属分类单位时,植物与细菌和动物的命名法规不同。《国际植物命名法规》规定,植物的亚属(组)名必须在属名和亚属名之间加等级指定词[亚属(subg.)、组(sect.)],亚属名用斜体,首字母大写,等级指定词用正体^[12],如 *Costus* (属) subg. *Metacostus* (亚属)。《国际细菌命名法规》《国际动物命名法规》规定,细菌和动物的亚属名必须置于圆括号内插入在属名和种本名之间,亚属名用斜体,首字母大写^[13-14],如 *Bacillus* (*Acrobacillus*) *polymyxa* (细菌)、*Aus* (*Bus*) *intermedius* (动物)。如果将亚属名的首字母用小写,就会被误认为是一个种本名,表示一个属级类群单元内若干个种的集合^[14],如 *Ornithoptera* (*priamus*) *lydius* 表示 *Ornithoptera priamus* 集合的 *lydius* 成员,这点需注意区别。

存在亚种分类单位时,动物与细菌和植物的命名法规不同。《国际动物命名法规》规定,动物的亚种名由属名、种名、亚种名组成,亚种本名以小写字母开始,用斜体^[14],如 *Arvicola* (属) *terrestris* (种) *pallasi* (亚种)。《国际细菌命名法规》《国际植物命名法规》规定,细菌和植物的亚种名由属名、种加词、亚种一词的缩写词(subsp.、ssp.、sspp.)及亚种名组成,缩写词用正体,其他同动物^[12-13],如 *Bacillus cereus* subsp. *mycoides* (细菌)、*Andropogon ternatus* subsp. *macrothrix* (植物);种之下的其他分类单位,包括种(sp.、spp.)、变种(var.)、亚变种(subvar.)、变型(f.)、亚变型(subf.)等的编排

同亚种,如 *Saxifraga aizoon* var. *aizoon* subvar. *brevifolia* f. *multicaulissubf. surculosa*^[12-13]。切勿将等级指定词编排成斜体(*Andropogon ternatus* subsp. *macrothrix*),将种以下分类单位编排成正体(*Bambusa dissimulata* McClure var. *albinodia* McClure^[15]),或将首字母编排成大写(*Andropogon ternatus* subsp. *Macrothrix*)。

1.3 栽培品种

《国际栽培植物命名法规》对栽培植物的名称进行了规范,基本形式是属名、种加词、品种加词,也可以写出种下分类群名称。品种加词可以用任何语言的一个或几个词表示,放在单引号内;每个词的首字母一般大写,用正体,但不位于首位的连词、介词、冠词和特殊的词或词组除外;以汉字书写的品种加词没有大小写的区别,但以汉语拼音表示时,也应遵守大小写的规则,如, *Cyclamen hederifolium* 'Artemis'、*Cyclamen hederifolium* var. *hederifolium* 'Artemis'、*Narcissus* 'Golden Harvest'、*Paeonia suffruticosa* '乌龙捧盛'等^[16-17]。如果品种加词不放入单引号内,会被误认为是命名人名称,这点需注意。

2 基因

基因工程的主要目的是从供体生物获取具有特定功能的基因。用反映基因的某些性状的缩写符号命名基因,并用斜体表示^[18-19],已被学术界接受,也推广到了所有物种。由于自然界物种多样,基因特点多样,又逐渐形成了以罗马字母的大小写、上下角标等来命名基因以区别物种属性、基因类别等的约定,对此相关文献进行了分析,但对基因符号中的数字及表示基因特性的符号的编排有不同的理解。

2.1 正斜体

基因符号中的字母用斜体已毋庸置疑,但对字母后的数字编排成正体还是斜体有不同的看法。有的文献明确表示阿拉伯数字用正体^[20];有的文献虽没明确提出阿拉伯数字用正体,但编排的结果,如 *ade5*、*his2-1*、*Dek12-W23*^[6],以及 *lacY1*、*lac-23*、*Mc1r*^[21]无不是正体;有的文献明确提出阿拉伯数字和罗马数字用斜体^[22]。《TIG 遗传命名指南》中文版对某些由字母和数字组成的基因符号表述为“由三个小写斜体字母与一个阿拉伯数字组成”(如裂殖酵母、玉米的基因符号),仅表明字母用斜体,并未明确说明阿拉伯数字用斜体还是用正体,但其所列物种基因符号字母后的阿拉伯数字和罗马数字均为斜体,如 *cde25*、*spoIIB*,包括排在下角的阿拉伯数字也用斜体,如 *glnAp₁*、*glnAp₂*。

对将表示基因属性、类别标于主符号上或下角的字母编排成正体还是斜体没有相应的规范遵循,如将

四环素抗性基因中表示抗性的“r”编排成 tet^r 或 tet^r 。在《TIG 遗传命名指南》中文版中,表示基因属性的上、下角均为斜体,如将热敏感“ts”和冷敏感“cs”的基因表示成 $cdc2-5^{ts}$ 、 $top2-250^{cs}$,将物种或株系中同名的基因表示成 $hisA_{E. coli}$ 。

从《TIG 遗传命名指南》是目前国际上同行业中最权威、应用最广泛的分子生物学命名规则^[23],也是国内较权威的期刊社及出版社如《遗传学报》、科学出版社等都执行的规则^[6],而且国外 SCI 收录期刊基因符号的字母和数字为斜体的期刊占 90.00%^[22]来看,本文赞同将基因(包括假基因)符号的所有组成部分,包括阿拉伯数字、罗马数字、表示基因属性的上下角均编排成斜体。

2.2 大小写

《TIG 遗传命名指南》对不同物种基因符号字母的大小写有不同的约定。例如,将微生物中的细菌,以及裂殖酵母、酿酒酵母等真菌用小写字母表示,如 pur 、 $arg1$ 、 $cdc28$;而细菌病毒噬菌体基因符号的字母大、小写均可,如 N 、 int 。植物依据物种不同,基因符号字母的大小写也不同,玉米用小写字母,如 $adh1$;拟南芥用大写字母,如 EMB ;而豌豆大、小写均可,如 Np 、 lam 。动物中大鼠和小鼠的基因符号通常首字母大写,如 $Tytp$ 、 $Mclr$;而人类的基因符号字母均用大写,如 $KRT1$ ^[19]。

本文认为,遇到同一基因符号字母大小写不一致时,首先分析此基因的物种来源,然后根据《TIG 遗传命名指南》对不同物种基因符号中字母大小写的约定确定。例如,四环素抗性基因就有 tet^r 、 Tet^r 、 TET^r 等形式,由于四环素抗性基因来源于大肠杆菌,而大肠杆菌是细菌,根据《TIG 遗传命名指南》的约定,细菌基因符号的字母用小写,应将四环素抗性基因编排成 tet^r ,判断其他基因符号的大小写同此。

3 工具酶

从供体生物获取功能基因后,需要对其进行切割、连接、聚合等操作,以限制性内切核酸酶、DNA 连接酶和 DNA 聚合酶为主的多种工具酶的应用,为基因操作提供了十分重要的技术基础。文稿中对它们的编排也有一定的要求。

限制性内切核酸酶的作用是将 DNA 切割成大小不同的片段,对其编排的规则如下:用第一次提取到此酶的生物体属名的第一个大写字母(如属名 *Haemophilus* 的 H)与种名的前 2 个小写字母(如种名 *influenzae* 的 in)组成的 3 个斜体字母(Hin)表示酶的基本名称;如果此酶存在于某一特殊的菌株中(如 Rd 菌

株),则将菌株名的一个字母(如 d)加在基本名称后,用正体表示(如 $Hind$);如果编码此酶的基因位于噬菌体或质粒上,则用一个正体大写字母表示这些非染色体的遗传因子,如 $EcoR$ 表示编码 $EcoR$ 的基因位于 *Escherichia coli* 中的抗药性 R 质粒上;用正体罗马数字表示在生物体中发现此酶的先后次序,如 $HindI$ 、 $HindII$ 、 $EcoRI$ 。以上是 Smith H O 和 Nathans D 于 1973 年提议的根据酶来源的生物体名称来命名限制性内切酶的规则^[24],已被学术界接受,但要注意字母与罗马数字之间不留空^[25]。

切割后的 DNA 片段须连接起来组成新的杂种 DNA 片段,这需要 DNA 连接酶的作用。最常用的 DNA 连接酶有 T4 DNA 连接酶和 T7 DNA 连接酶,但发现对 T4 和 T7 的编排有正体、斜体、数字下角等形式,如 $T4$ 、 $T4$ 、 T_4 。因为 T4 DNA 连接酶是病毒肠杆菌 T4 噬菌体 DNA 编码的,通过查阅与病毒命名相关的文献^[8-10]发现,病毒名称中的阿拉伯数字均为平排,加之病毒名称仅在指代分类系统中的正式种名时使用斜体,在一般性叙述中不必使用斜体^[5]。本文认为应将 T4 编排成正体、4 平排,即 T4 DNA 连接酶, T7 的编排同 T4。

不同 DNA 片段连接在一起后还需要一类修复 DNA 损伤和催化 DNA 复制的酶,即 DNA 聚合酶,主要有 T4 DNA 聚合酶、T7 DNA 聚合酶、 Taq DNA 聚合酶等。T4 DNA 聚合酶和 T7 DNA 聚合酶的编排规则如 T4 DNA 连接酶和 T7 DNA 连接酶。文章中常见将 Taq DNA 聚合酶中的 Taq 编排成正体,但根据 Taq DNA 聚合酶来源于水生栖热菌(*Thermus aquaticus*), Taq 取其拉丁名属名的首字母 T 和种名前 2 个字母 aq ,应编排成斜体,文献^[26]对此进行了详细的分析。

4 载体

基因工程的本质是使功能基因在特定条件下得到扩增和表达,但功能基因自身无法复制和表达,也不易进入受体细胞,这就需要一种能携带功能基因进入受体细胞,并使之在受体细胞中得以维持或表达的 DNA 分子,即载体。常用的载体有质粒载体、病毒载体等,因病毒的缩写一般编排成正体、平排^[5],病毒载体的编排也为正体、平排,如细菌病毒噬菌体载体 λ 噬菌体载体、M13 噬菌体载体等。

质粒载体是基因工程最常用的载体之一,其编排规则为:天然存在的质粒用原有的命名^[19],构成质粒命名的字母与数字均编排为正体;新发现或改造的质粒命名由小写的前缀“p”(plasmid 的首字母)加非斜体字母(发现或构建该质粒的作者或实验室名称)和

编号(该实验室给该质粒的顺序编号)组成^[19,27]。例如,质粒 pBR322 是由 3 个天然质粒 pSC101(含有抗生素的抗性质粒,即 R 质粒)、ColE1(产大肠杆菌素的 Col 质粒)、pSF2124(R 质粒)构建而成,它们均编排成正体。另外,作为载体的质粒一般来源于细菌或真菌,因为细菌和真菌的基因符号用小写编排,本文认为质粒上的基因符号也应用小写编排;当质粒缺失基因或插入外源基因,质粒的命名可加上相应的缺失基因或外源基因,将基因符号编排成斜体,如 pI258Δ7(*cad-asa*)(缺失了基因 *cad* 与 *asa*)^[19,27]、重组质粒 pcDNA3-*dnaJ*。

5 结束语

优秀的学术文稿不仅体现在内容的科学性、准确性上,还体现在符号语言编排的规范性上,尤其在基因工程文稿中,相同符号的正斜体、大小写代表的含义均不同,因此对符号的编排有更高的要求。本文建议基因供体和受体生物的符号按照《国际病毒分类与命名原则》《国际细菌命名法规》《国际藻类、真菌、植物命名法规》《国际栽培植物命名法规》《国际动物命名法规》的要求进行编排;基因符号按照《TIG 遗传命名指南》的要求进行编排;工具酶中的限制性内切酶和 *Taq* 酶按照酶来源的生物体名称编排规则进行编排;载体上的基因根据载体来源,按照《TIG 遗传命名指南》的要求进行编排。

6 参考文献

- [1] 全国科学技术名词审定委员会. 遗传学名词[M]. 2 版. 北京: 科学出版社, 2006
- [2] 杜秀杰, 赵大良. 学术论文语言表达范式分析[J]. 编辑学报, 2018, 30(3): 260
- [3] 夏元铃, 高珏, 周丽萍, 等. 分子生物学类论文中几类常见名词的正斜体编排[J]. 编辑学报, 2009, 21(2): 127
- [4] 高艳. 国家标准审编基础: 要素表述(五): 生命科学符号中物种名称的表示[J]. 中国质量与标准导报, 2018(12): 26
- [5] 汪继祥. 科学出版社作者编辑手册[M]. 北京: 科学出版社, 2004
- [6] 张冰. 科技期刊中基因及蛋白质的规范表达[G]//赵惠祥. 学报编辑论丛: 2007. 上海: 上海大学出版社, 2007: 81
- [7] 王宇, 李段, 张乐宜, 等. 猪伪狂犬病毒的分离鉴定及其 gE、TK 全基因序列遗传变异分析[J]. 广东农业科学, 2016, 43(7): 133
- [8] MAYO M A, HORZINEK M C. A revised version of the international code of virus classification and nomenclature [J]. Archives of Virology, 1998, 143(8): 1645
- [9] 徐耀先, 解梦霞, 向近敏. 病毒命名与分类系统研究进展[J]. 中国病毒学, 1999, 14(3): 190
- [10] 洪健, 周雪平. ICTV 第九次报告以来的植物病毒分类系统[J]. 植物病理学报, 2014, 44(6): 561
- [11] 张忠信. ICTV 第九次报告对病毒分类系统的一些修改[J]. 病毒学报, 2012, 28(5): 595
- [12] MCNEILL J, BARRIE F R, BURDET H M, 等. 国际植物命名法规(维也纳法规)[M]. 张丽兵, 译. 北京: 科学出版社, 2007
- [13] LAPAGE S P, SNEATH P H A, LESSEL E F, 等. 国际细菌命名法规[M]. 陶天申, 陈文新, 骆传好, 译. 北京: 科学出版社, 1989
- [14] RIDE W D L, COGGER H G, DUPUIS C, 等. 国际动物命名法规[M]. 卜文俊, 郑乐怡, 译. 4 版. 北京: 科学出版社, 2007
- [15] 史军义, 易同培. 国际两大植物命名体系及其相互关系[J]. 生物学通报, 2017, 52(1): 5
- [16] 布里克尔, 亚历山大, 戴维, 等. 国际栽培植物命名法规[M]. 靳晓白, 成仿云, 张启翔, 译. 8 版. 北京: 中国林业出版社, 2013
- [17] 靳晓白. 栽培植物命名法与国际栽培品种登录[J]. 农业科技与信息(现代园林), 2015, 12(8): 584
- [18] DEMEREC M, ADELBERG E A, CLARK A J, et al. A proposal for a uniform nomenclature in bacterial genetics [J]. Genetics, 1966, 54(1): 61
- [19] 《遗传学进展》编辑部. TIG 遗传命名指南[M]. 王金发, 陈中健, 杨琳, 等译. 北京: 科学出版社, 2002
- [20] 原丽欣. 分子生物学领域几种常见名词符号的规范编排[J]. 中国科技期刊研究, 2011, 22(6): 962
- [21] 张翠英. 基因及蛋白质符号的规范编排[J]. 编辑学报, 2004, 16(4): 262
- [22] 周桂莲, 许育彬, 席美丽, 等. 也谈科技期刊中几种常见分子生物学名词符号的规范编排[J]. 天津科技, 2018, 45(2): 78
- [23] 吴敬涛, 杨桂英. 谈科技论文中分子生物学符号的编辑排版方法[J]. 新闻传播, 2013(9): 70
- [24] SMITH H O, NATHANS D. A suggested nomenclature for bacterial host modification and restriction systems and their enzymes [J]. Journal of Molecular Biology, 1973, 81(3): 419
- [25] 郭建顺, 沈晓峰, 张学东, 等. 分子生物学书刊中限制性内切酶的规范表达[J]. 编辑学报, 2005, 17(3): 195
- [26] 王连芬, 张立方, 孙勇. *Taq* 酶正斜体编排问题的探讨[J]. 中国科技期刊研究, 2011, 22(2): 291
- [27] NOVICK R P, CLOWES R C, COHEN S N, et al. Uniform nomenclature for bacterial plasmids: a proposal [J]. Bacteriological Reviews, 1976, 40(1): 168

(2021-03-10收稿;2021-05-23修回)