组分距离方法构建基于核糖体蛋白质或氨酰tRNA合成酶的原核生物亲缘树*

卫海滨[①②†] 戚 继[③④†] 郝柏林[②③**]

(① 浙江大学生命科学学院, 杭州 310027; ② 中国科学院北京基因组研究所杭州分部, 杭州 310008; ③ 复旦大学理论生命科学研究中心, 上海 200433; ④ 中国科学院理论物理研究所, 北京 100080)

摘要 最近新发展的基于计算短串频度, 推断原核生物系统发生关系的 *K* 串组分距离方法, 以整个蛋白质组作为数据集. 采用每个物种的全部核糖体蛋白质或氨酰 tRNA 合成酶, 得到一致结果. 已知后一组蛋白质在单个使用时, 产生彼此不同的进化树. 我们构建进化树不需做任何序列联配, 所得亲缘树包括 16 个古细菌、105 个细菌和 2 个真核生物. 大部分低层分支与《伯杰系统细菌学手册》(第 2 版), 即 2003 年第 4 次发布的细菌系统分类大纲相一致, 而且对高层分支关系给出一些建议.

关键词 原核生物 古细菌 系统发生 系统发生树 组分距离

原核生物的系统分类因缺乏形态学特征, 而成为微生物学中的一大挑战[1]. 20 世纪 70 年代, Woese 等人[2]通过联配核糖体小亚基RNA(SSU rRNA)序列推断系统发生关系, 取得突破性进展. 他们发现生命领域中除细菌和真核生物外还有第三个超界: 古细菌, 而且支持了叶绿体和线粒体的内共生起源学说. 这是rRNA构树研究的主要成就. 核糖体rRNA数据库的建立[3,4]使得人们易于构建基于SSU rRNA的分子系统发生树. 以至在《伯杰系统细菌学手册(简称伯杰手册)》(第 2 版)中写到: “更多的基于分析SSU rRNA核苷酸序列而非表型特征所构建的系统发生框架”(参看George Garrity的序言[5]).

然而, 近年来对单独使用SSU rRNA推断系统发生关系的可靠性产生了质疑. 这些大约有 1500 bp的核苷酸序列, 不可能提供足够的系统发生信息来分辨生命之树的所有分支, 甚至有证据表明这些保守的RNA也可能存在水平转移[6,7]. 此外, 自 1995 年开始不断增加的原核生物完全基因组, 对分子系统发生关系的分析带来更多的问题. 因而现在的一般看法是, 不同的基因反映不同的历史, 基因树不能等同

2003-11-24 收稿, 2003-12-18 收修改稿

* 国家重点基础研究项目(批准号: G2000077308)、国家自然科学基金(批准号: 30170232)和中国科学院创新工程资助项目及上海市通过复旦大学的资助

** 联系人, E-mail: hao@itp.ac.cn

† 同等贡献

于物种树. 特别是在分子系统发生学上, 基因横向传递和世系之间保守基因的丢失已经成为讨论的热点[8,9]. 为了充分利用不断增加的基因组数据, 最近发展出许多基于全基因组的方法[10,11]. 介于使用单基因和全基因组两种极端之间, 还提出了基于联合蛋白质序列的方法[12]. 不过, 所有这些方法在不同阶段或明或暗地需要序列联配和打分方案, 因此依赖于许多参数和微调.

为了避免序列联配和对基因的特殊选择, 我们发展了一种基于K串的组分距离方法, 从完全基因组出发推断系统发生关系. 该方法已被成功地应用于原核生物[13]、叶绿体[14]和冠状病毒[15]的进化研究. 然而, 使用完全基因组数据同时也可以视为这种方法的缺点. 因此, 文中我们采用两组蛋白质, 它们包含的系统发生信息可能十分不同. 核糖体蛋白质和rRNA交织成复合大分子, 作为一个整体来行使功能, 因此它们不易与别的物种发生横向传递. 用连接核糖体蛋白质序列的方法很自然地产生合理的系统发生关系[10,16]. 相反, 氨酰tRNA合成酶作为独立的分子行使功能, 在不同物种之间发生横向传递没有很大障碍. 事实上, 人们已经知道它不适合用于系统发生推断. 如果单独使用 20 种不同的氨酰tRNA合成酶产生 20 种不同的树, 有一些甚至不能区分开生命之树上古细菌、细菌和真核生物三个超界[17~19]. 然而我们的结果表明收集一个物种所有的氨酰tRNA合成酶可以产生与核糖体蛋白质树和全蛋白质组树[13]相一致的系统发生树.

本文的目的有三个方面. 第一, 说明组分距离方法不必要求完全蛋白质组数据, 选取一个合适家族的蛋白质序列也可行. 第二, 提供一种新的分子系统发生分析方法, 它独立于但支持基于 SSU rRNA 序列的“标准”方法. 第三, 代替仅仅使用自举法或刀切法的稳定性和一致性测试, 通过与细菌分类学结果的严格比较, 验证这种新方法.

1　材料和方法

在NCBI数据库中存在两套原核生物基因组, 一套位于Ganbank目录[20]下, 是作者提交的原始序列; 另一套经NCBI人员[21]重新整理和注释过, 索取号带有前缀NC_以示区别. 我们使用后一套中所有的原核生物完全基因组, 截止日期到 2003 年 6 月 10 日; 除去一个物种*Pasteurella multocida*, 因为在它的注释中没有核糖体蛋白质和氨酰tRNA合成酶的信息. 附录中给出物种名字、缩写、NCBI索取号以及在《伯杰手册》中的位置.

我们采用基于K串的组分矢量方法计算距离矩阵, 见文献[13], 因此, 下面只给出方法的简洁概述. 第一步, 从一个蛋白质家族或者完全基因组中收集所有的氨基酸序列. 第二步, 计算定长为K的寡肽出现的频度. 为了减弱在分子水平上随机中性突变的影响和突出选择进化的后果, 用$(K-2)$阶的Markov模型从这些频度中减除随机背景. 第三步, 将这些规范处理后的频度按固定次序排放, 为每个物种形成一个 20^K维的组分矢量. 第四步, 物种A和B的相关性$C(A, B)$由两者的组分矢量之间的夹角余弦决定. 因此, 如果两个矢量相同, 则相关性最高, $C = 1$; 如果它们没有共同的组分, 则相关性$C = 0$, 两个矢量成正交关系. 最后, 两个物种之间的距离可以定义为$D = (1 - C)/2$. 一旦获得了距离矩阵, 则采用标准方法[22], 如Phylip软件包[23]中的邻接算法来构建亲缘树. 随着K值的增加和对蛋白质序列再抽样检验, 树的拓扑结构也趋于稳定. 关于统计检验和方法的更多介绍参看文献[13,14].

2　结果和讨论

图 1 为基于核糖体蛋白质的亲缘树, 基于氨酰tRNA合成酶的亲缘树见图2. 计算包括所有123个物种, 然而由于相同种的不同菌株和同一属下不同的种总是形成一组, 因此在最后的图中仅仅保留了每个属下的一个物种作为代表. 因此, 图 1 和 2 相当于属的进化树.

我们的亲缘树描述了 121 个原核生物, 涵盖了 67 个属, 55 个科, 46 个目, 25 个纲和 25 个原核生物门中的 14 个门, 因而可以与细菌学家的分类做一个详尽且更严格的比较. 事实上, 我们要和 3 种不同而又相关的树作比较: SSU rRNA树[1], 它是一棵合成树, 包含 253 个物种; RDP-Ⅱ骨干树 8.0 版[4], 包括 217 个序列, 代表《伯杰手册》(第 2 版)的 203 个科中的 187

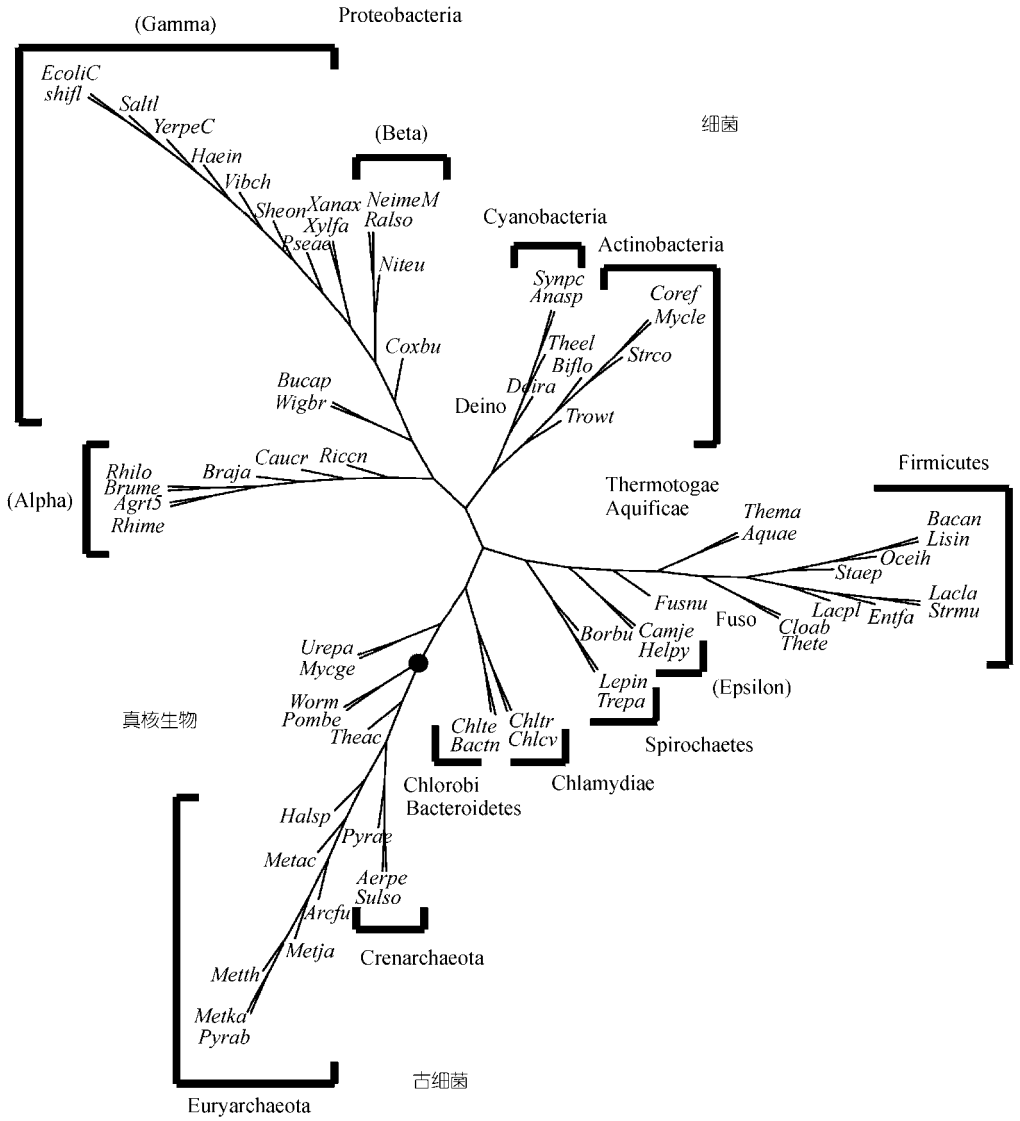

图 1 基于核糖体蛋白质的原核生物分子亲缘树

计算使用的字符串长度为 $K = 5$. 图中所示为 14 个门中的 67 个属, 每个门的名字放在靠近物种的位置. 对于具有最多已测基因组数据的变形菌 B12 门 Proteobacteria, 我们将其下属的纲也标识在括弧中. 这是一棵无根树, 分支长度并没有按比例表示. 图中黑色圆点标识生命之树中三个超界的分叉点

个; 最后是《伯杰手册》本身[5,24], 它主要基于SSU rRNA模型, 但是也考虑了传统的分类.

一般来说, 基于核糖体蛋白质的亲缘树和SSU rRNA树的一致性较高, 而基于氨酰tRNA合成酶的亲缘树与SSU rRNA树的一致性较低[17~19]. 然而, 后者在本文结果中的表现远好于基于单个氨酰tRNA合成酶的树. 特别是, 两组进化树都划分出生命领域的三大超界, 这是它们一致的显著特征.

由于属到科、到目的分支基本上与 SSU rRNA 树一致, 因此我们集中讨论在不同分类水平上的差异, 尤其是可能对分类提出修正的显著差异.

种间的详细位置关系在属树上看不到. RDP-Ⅱ骨干树[4]和我们的进化树(图 1 和 2)都是如此, 但是有两个细节反映在我们的更详细的物种树上. 第一, 解

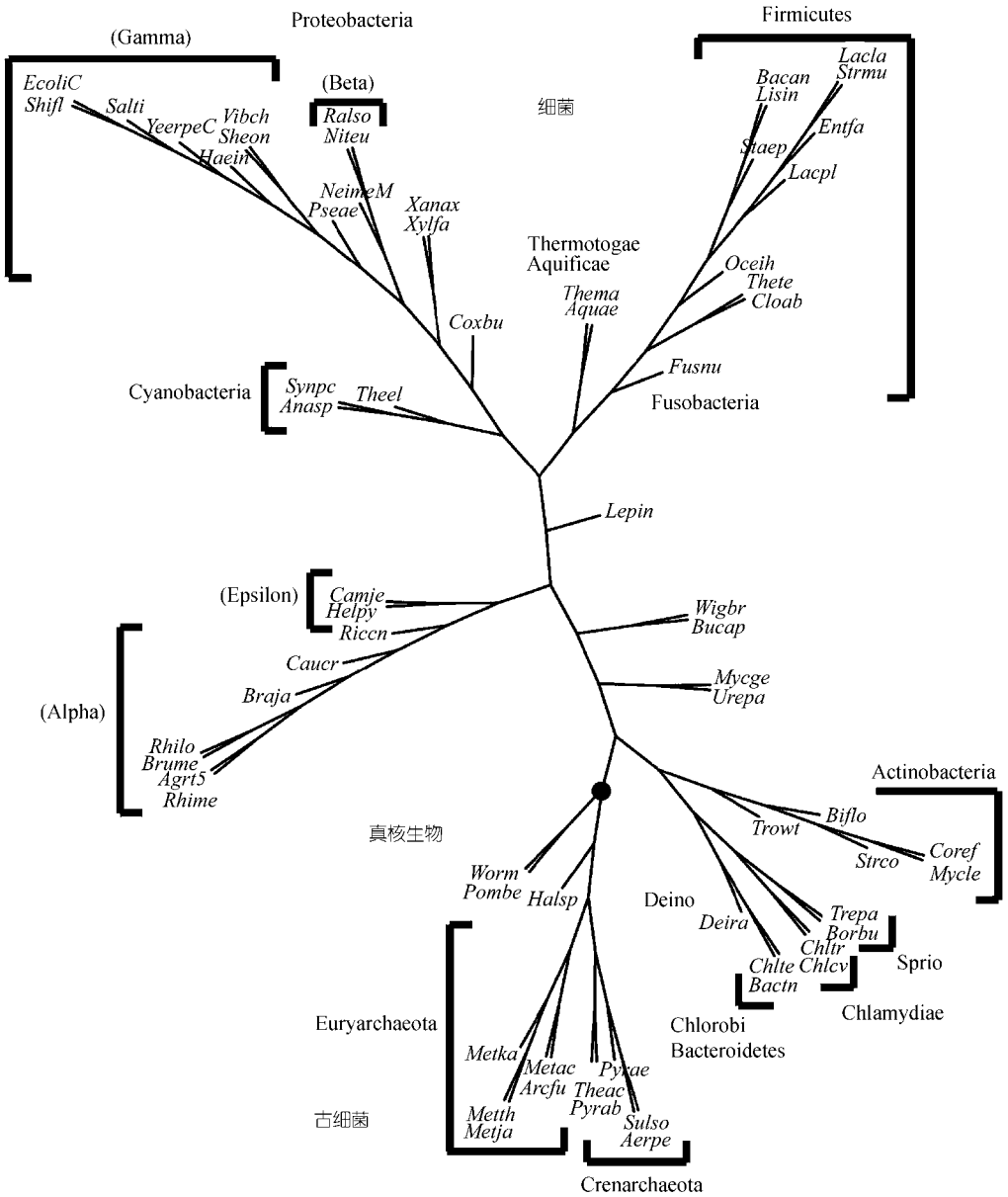

图 2　由氨酰 tRNA 合成酶集合构建的原核生物分子系统发生树

计算使用的字符串长度为 $K = 5$. 其余部分同图 1

脲支原体Urepa插入到另一个支原体属*Mycoplasma*中，在SSU rRNA树[21]中也存在同样的现象．第二，志贺氏痢疾杆菌Shifl插入到大肠杆菌属*Escherichia*中．后者尚有待于新的SSU rRNA树的验证．在高层分类水平上，SSU rRNA树[21]中变形菌*Proteobacteria*的Beta纲插在Gamma纲中，这同样反映在本文和文献[13]的亲缘树上．进一步观察会发现Gamma纲被Beta纲隔开的两个属(*Buchnera*和*Wigglesworthia*)拥有较小的基因组，后者在氨酰tRNA合成酶树(图 2)上远离主要的Proteobacteria门．在所有这些树中具有较小基因组的物种，形成处于更深层次的亚组．小基因组自然应该进化更早，这可能是真实进化史的表现．总之，基

因组大小所带来的影响不易清楚地表现在过去基于单基因的进化树上.

所有 3 个螺旋菌(Spirochetes, *Burbu*, *Trepa*和*Lepin*), 虽然在图 1 和《伯杰手册》中都被归为一组, 但是在图 2 和蛋白质组树[13]上, *Lepin*却远离螺旋菌门. *Lepin*的基因组比另外两个大很多, 不过, 我们仍然不能断定这是否由基因组大小的显著差异所造成.

古细菌*Methanopyrus kandleri* (Metka)曾在SSU rRNA进化树分析中, 被预测远离其他的产甲烷古细菌[25]. 但在图 1 和 2 中它均和那些产甲烷菌形成一个分支; 在它的基因组报告[26]中, 基因成分和基因对的分析结果也表明如此. 另外, Crenarchaeota门所含的 3 个属: *Pyrae*, *Aerpe*和*Sulso*总是处于一个分支. 但是Euryarchaeota门所含的*Halsp*和*Theac*所处位置不稳定, 有时在别的树上也观察到类似行为[10,11].

以Oceih为代表的新属*Oceanobacillus*, 曾是惟一一个跨门的差别. 在图 1 和 2 与$K = 5$, $K = 6$ 的蛋白质组树[13]上, 它位于B13 门(Firmicutes), 与*Bacillus*属(Bergey's代码为B13.3.1.1)的位置较近. 这同NCBI[21]分类一致, 但是在 2002 年《伯杰手册概要》[24]中却将它分入*Gammaproteobacteria* (B12.3.8.1.6), 同时带着这样一个脚注: "*Oceanospirillales*在ARB树上的位置不明确". 然而, 2003 年发表的《伯杰手册概要》[27]将Oceih移动到B13.3.1.1.12, 相应地, 附录表 2 中Oceih也已调整到它的正确位置.

我们将简要地提一下高层分支问题. 高层分支的划分和定位远非原核生物而是一般分类上的常见争论问题. 在诸如《伯杰手册》的分类系统中, 许多门并列于古细菌和细菌超界之下. 将文献[13]和本文的进化树与SSU rRNA树[1], RDP-Ⅱ骨干树[4], 与别的全基因组方法[10,11]获得的进化树相比较, 能够识别出一批这些进化树在高层分支上的共同特征, 因而它们不像是偶然的假象. (ⅰ) 两个门Aquificae (B1)和Thermotogae (B2)总是紧密相连, 形成一个分支, 然后再加入进化树的主干. (ⅱ) Chlorobi (B11)和Bacteroides (B20)门先聚成一个分支, 在文献[1]和[19]中也是如此. (ⅲ) Chlamydiae (B16)门和Spirochaetes (B17)门连接到进化树的分支点总是相距很近(除了一些进化树上Lepin远离B17 的例外). (ⅳ) Deinococcus-Thermus (B4)和Actinobacteria (B14)在许多进化树上出现在一起. (ⅴ) *Mycoplasma*属独立于Firmicutes (B13)门, 形成单独的分支, 在许多全基因组树包括我们的进化树上, 这都是一个显著特征.

然而, 应该注意到目前 14 个门中有 6 个门仅有一个代表物种, 以后有广阔分类背景的基因组数据时, 高层分支的关系将进一步得到验证. 组分距离方法作为一种新的推断系统发生关系的方法, 无需序列联配和参数调节, 它和传统的SSU rRNA[1]分析一起应能给出原核生物分类的统一的分子基础.

附录: 我们的工作中使用了 16 个古细菌、105 个细菌和 2 个真核生物的数据, 所有物种的名字、缩写和索取号列于下面的表 1~3 中. 表 1 和 2 的最后一列的Berger's 代码, 是《伯杰细菌系统学手册》(第 2 版)的分类缩写. 例如, EcoliK 属于门 BXII (Proteobacteria)、纲III (Gammaproteobacteria)、目XIII (Enterobacteriales)、科Ⅰ(Enterobacteriaceae)和属Ⅰ(*Echerichia*). 我们把罗马数字换成阿拉伯数字, 去掉分类中的拉丁名字, 将它缩写为 B12.3.13.1.1.

表 1 16 个古细菌的名字、缩写和 NCBI 索取号(按 Bergey's 代码排序)

古细菌名字	缩写	索取号	Bergey's 代码
Pyrobaculum aerophilum	Pyrae	NC_003364	A1.1.1.1.1
Aeropyrum pernix K1	Aerpe	NC_000854	A1.1.2.1.3
Sulfolobus solfataricus	Sulso	NC_002754	A1.1.3.1.1
Sulfolobus tokodaii	Sulto	NC_003106	A1.1.3.1.1
Methanobacterium thermoautotrophicus	Metth	NC_000916	A2.1.1.1.1
Methanococcus jannaschii	Metja	NC_000909	A2.2.1.1.1
Methanosarcina acetivorans str. C2A	Metac	NC_003552	A2.2.3.1.1
Methanosarcina mazei Goel	Metma	NC_003901	A2.2.3.1.1

续表 1

古细菌名字	缩写	索取号	Bergey's 代码
Halobacterium sp. NRC-1	Halsp	NC_002607	A2.3.1.1.1
Thermoplasma acidophilum	Theac	NC_002578	A2.4.1.1.1
Thermoplasma volcanium	Thevo	NC_002689	A2.4.1.1.1
Pyrococcus abyssi	Pyrab	NC_000868	A2.5.1.1.3
Pyrococcus furiosus	Pyrfu	NC_003413	A2.5.1.1.3
Pyrococcus horikoshii	Pyrho	NC_000961	A2.5.1.1.3
Archaeoglobus fulgidus	Arcfu	NC_000917	A2.6.1.1.1
Methanopyrus kandleri AV19	Metka	NC_003551	A2.7.1.1.1

表 2　105 个细菌的名字、缩写和NCBI索取号(按Bergey's代码排序)[a)]

细菌名字	缩写	索取号	Bergey's 代码
Aquifex aeolicus	Aquae	NC_000918	B1.1.1.1.1
Thermotoga maritima	Thema	NC_000853	B2.1.1.1.1
Deinococcus radiodurans R1	Deira	NC_001263-64	B4.1.1.1.1
Thermosynechococcus elongatus BP-1	Theel	NC_004113	B10.1.[a)]
Synechocystis PCC6803	Synpc	NC_000911	B10.1.1.1.14
Nostoc sp. PCC7120	Anasp	NC_003272	B10.1.4.1.8
Chlorobium tepidum TLS	Chlte	NC_002932	B11.1.1.1.1
Rickettsia conorii	Riccn	NC_003103	B12.1.2.1.1
Rickettsia prowazekii	Ricpr	NC_000963	B12.1.2.1.1
Caulobacter crescentus	Caucr	NC_002696	B12.1.5.1.1
Agrobacterium tumefaciens C58 Cereon	Agrt5	NC_003062-63	B12.1.6.1.2
Agrobacterium tumefaciens C58 UWash	Agrt5W	NC_003304-05	B12.1.6.1.2
Sinorhizobium meliloti 1021	Rhime	NC_003047	B12.1.6.1.6
Brucella melitensis	Brume	NC_003317-18	B12.1.6.3.1
Brucella suis 1330	Brusu	NC_004310-11	B12.1.6.3.1
Mesorhizobium loti	Rhilo	NC_002678	B12.1.6.4.6
Bradyrhizobium japonicum	Braja	NC_004463	B12.1.6.7.1
Ralstonia solanacearum	Ralso	NC_003295-96	B12.2.1.2.1
Neisseria meningitidis MC58	NeimeM	NC_003112	B12.2.4.1.1
Neisseria meningitidis Z2491	NeimeZ	NC_003116	B12.2.4.1.1
Nitrosomonas europaea ATCC 19718	Niteu	NC_004757	B12.2.5.1.1
Xanthomonas axonopodis citri 306	Xanax	NC_003919	B12.3.3.1.1
Xanthomonas campestris ATCC 33913	Xanca	NC_003902	B12.3.3.1.1
Xylella fastidiosa	Xylfa	NC_002488	B12.3.3.1.9
Xylella fastidiosa Temecula1	Xylft	NC_004556	B12.3.3.1.9
Coxiella burnetii RSA 493	Coxbu	NC_002971	B12.3.6.2.1
Pseudomonas aeruginosa PA01	Pseae	NC_002516	B12.3.9.1.1
Pseudomonas putida KT2440	Psepu	NC_002947	B12.3.9.1.1
Pseudomonas syringae pv *tomato* str.DC3000	Psesy	NC_004578	B12.3.9.1.1
Shewanella oneidensis MR-1	Sheon	NC_004347	B12.3.10.1.7
Vibrio cholerae	Vibch	NC_002505-06	B12.3.11.1.1
Vibrio parahaemolyticus RIMD 2210633	Vibpa	NC_004603.05	B12.3.11.1.1
Vibrio vulnificus CMCP6	Vibvu	NC_004459-60	B12.3.11.1.1
Buchnera aphidicola Sg	Bucap	NC_004061	B12.3.13.1.5
Buchnera aphidicola (*Baizongia pistaciae*)	Bucbp	NC_004545	B12.3.13.1.5
Buchnera sp. APS	Bucai	NC_002528	B12.3.13.1.5
Escherichia coli CFT073	EcoliC	NC_004431	B12.3.13.1.13
Escherichia coli K12	EcoliK	NC_000913	B12.3.13.1.13
Escherichia coli O157:H7	EcoliO	NC_002695	B12.3.13.1.13

续表 2

细菌名字	缩写	索取号	Bergey's 代码
Escherichia coli O157:H7 EDL933	EcoliE	NC_002655	B12.3.13.1.13
Salmonella typhi	Salti	NC_003198	B12.3.13.1.32
Salmonella typhi Ty2	SaltiT	NC_004631	B12.3.13.1.32
Salmonella typhimurium LT2	Salty	NC_003197	B12.3.13.1.32
Shigella flexneri 2a str. 2457T	Shifl2	NC_004741	B12.3.13.1.34
Shigella flexneri 2a str. 301	Shifl	NC_004337	B12.3.13.1.34
Wigglesworthia brevipalpis	Wigbr	NC_004344	B12.3.13.1.38
Yersinia pestis strain C092	YerpeC	NC_003143	B12.3.13.1.40
Yersinia pestis KIM	YerpeK	NC_004088	B12.3.13.1.40
Haemophilus influenzae Rd	Haein	NC_000907	B12.3.14.1.3
Campylobacter jejuni	Camje	NC_002613	B12.5.1.1.1
Helicobacter pylori 26695	Helpy	NC_000915	B12.5.1.2.1
Helicobacter pylori J99	Helpj	NC_000921	B12.5.1.2.1
Clostridium acetobutylicum ATCC824	Cloab	NC_003030	B13.1.1.1.1
Clostridium perfringens	Clope	NC_003366	B13.1.1.1.1
Clostridium tetani E88	Clote	NC_004557	B13.1.1.1.1
Thermoanaerobacter tengcongensis	Thete	NC_003869	B13.1.2.1.8
Mycoplasma genitalium	Mycge	NC_000908	B13.2.1.1.1
Mycoplasma penetrans	Mycpe	NC_004432	B13.2.1.1.1
Mycoplasma pneumoniae	Mycpn	NC_000912	B13.2.1.1.1
Mycoplasma pulmonis UAB CTIP	Mycpu	NC_002771	B13.2.1.1.1
Ureaplasma urealyticum	Urepa	NC_002162	B13.2.1.1.4
Bacillus anthracis str. Ames	Bacan	NC_003997	B13.3.1.1.1
Bacillus cereus ATCC 14579	Bacce	NC_004722	B13.3.1.1.1
Bacillus halodurans	Bachd	NC_002570	B13.3.1.1.1
Bacillus subtilis	Bacsu	NC_000964	B13.3.1.1.1
Oceanobacillus iheyensis	Oceih	NC_004193	B13.3.1.1.12
Listeria innocua	Lisin	NC_003212	B13.3.1.4.1
Listeria monocytogenes EGD-e	Lismo	NC_003210	B13.3.1.4.1
Staphylococcus aureus Mu50	StaauM	NC_002758	B13.3.1.5.1
Staphylococcus aureus MW2	StaauW	NC_003923	B13.3.1.5.1
Staphylococcus aureus N315	StaauN	NC_002745	B13.3.1.5.1
Staphylococcus epidermidis ATCC 12228	Staep	NC_004461	B13.3.1.5.1
Lactobacillus plantarum WCFS1	Lacpl	NC_004567	B13.3.2.1.1
Enterococcus faecalis V583	Entfa	NC_004668	B13.3.2.4.1
Streptococcus agalactiae 2603V/R	StragV	NC_004116	B13.3.2.6.1
Streptococcus agalactiae NEM316	StragN	NC_004368	B13.3.2.6.1
Streptococcus mutans UA159	Strmu	NC_004350	B13.3.2.6.1
Streptococcus pneumoniae R6	StrpnR	NC_003098	B13.3.2.6.1
Streptococcus pneumoniae TIGR4	StrpnT	NC_003028	B13.3.2.6.1
Streptococcus pyogenes MGAS315	StrpyG	NC_004070	B13.3.2.6.1
Streptococcus pyogenes MGAS8232	Strpy8	NC_003485	B13.3.2.6.1
Streptococcus pyogenes SF370	StrpyS	NC_002737	B13.3.2.6.1
Streptococcus pyogenes SSI-1	StrpyI	NC_004606	B13.3.2.6.1
Lactococcus lactis sp. IL1403	Lacla	NC_002662	B13.3.2.6.2
Corynebacterium efficiens YS-314	Coref	NC_004369	B14.(1.5).(1.7).1.1
Corynebacterium glutamicum	Corgl	NC_003450	B14.(1.5).(1.7).1.1
Mycobacterium leprae TN	Mycle	NC_002677	B14.(1.5).(1.7).4.1
Mycobacterium tuberculosis CDC1551	MyctuC	NC_002755	B14.(1.5).(1.7).4.1
Mycobacterium tuberculosis H37Rv	MyctuH	NC_000962	B14.(1.5).(1.7).4.1

续表 2

细菌名字	缩写	索取号	Bergey's 代码
Tropheryma whipplei TW08/27	Trow8	NC_004551	B14.(1.5).(1.9).6.3
Tropheryma whipplei Twist	Trowt	NC_004572	B14.(1.5).(1.9).6.3
Streptomyces avermitilis MA-4680	Straw	NC_003155	B14.(1.5).(1.11).1.1
Streptomyces coelicolor A3(2)	Strco	NC_003888	B14.(1.5).(1.11).1.1
Bifidobacterium longum NCC2705	Biflo	NC_004307	B14.(1.5).2.1.1
Chlamydia muridarum	Chlmu	NC_002620	B16.1.1.1.1
Chlamydia trachomatis	Chltr	NC_000117	B16.1.1.1.1
Chlamydophila caviae GPIC	Chlcv	NC_003361	B16.1.1.1.2
Chlamydophila pneumoniae AR39	ChlpnA	NC_002179	B16.1.1.1.2
Chlamydophila pneumoniae CWL029	ChlpnC	NC_000922	B16.1.1.1.2
Chlamydophila pneumoniae J138	ChlpnJ	NC_002491	B16.1.1.1.2
Borrelia burgdorferi	Borbu	NC_001318	B17.1.1.1.2
Treponema pallidum	Trepa	NC_000919	B17.1.1.1.9
Leptospira interrogans serovar lai str. 56601	Lepin	NC_004342-43	B17.1.1.3.2
Bacteroides thetaiotaomicron VPI-5482	Bactn	NC_004663	B20.1.1.1.1
Fusobacterium nucleatum ATCC 25586	Fusnu	NC_003454	B21.1.1.1.1

a)《伯杰手册》中并没有给出物种的完整世系

表 3　两个真核生物的名字、缩写和 NCBI 索取号

2 个真核生物	缩写	索取号
Schizosaccharomyces pombe	Pombe	NC_003421.23.24
Caenorhabitidis elegans	Worm	NC_003279-84

参　考　文　献

1 Olsen G J, Woese C R, Overbeek R. The winds of (evolutionary) change: breathing new life into microbiology. J Bacteriol, 1994, 176: 1~6

2 Woese C R, Fox G E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA, 1977, 74: 5088~5090

3 The European Ribosomal RNA database. Available at: http://oberon.fvms.ugent.be:8080/rRNA/index.html[DOI]

4 Cole J R, Chai B, Marsh T L, et al. The ribosomal database project (RDP-Ⅱ): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucl Acids Res, 2003, 31: 442~443. Available at: http://rdp.cme.msu.edu/pubs/NAR/Backbone_tree.pdf[DOI]

5 Bergey's Manual Trust. Bergey's Manual of Systematic Bacteriology. 2nd ed. New York: Springer-Verlag, 2001

6 Brocchieri L. Phylogenetic inferences from molecular sequences: review and critique. Theoretical Population Biology, 2001, 59: 27~40 [DOI]

7 Nomura M. Engineering of bacterial ribosomes: replacement of all seven *Escherichia coli* rRNA operons by a single plasmid-encoded operon. Proc Natl Acad Sci USA, 1999, 96(5): 1820~1822 [DOI]

8 Doolittle W F. Phylogenetic classification and the universal tree. Science, 1999, 284: 2124~2128

9 Ragan M A. Detection of lateral gene transfer among microbial genomes. Current Opinion in Genetics & Development, 2001, 11: 620~626[DOI]

10 Wolf Y I, Rogozin I B, Grishin N V, et al. Genome trees constructed using five different approaches suggest new major bacterial clades. BMC Evolutionary Biology, 2001, 1: 8. Available at: http://www.biomedcentral.com/1471-2148/1/8 [DOI]

11 Wolf Y I, Rogozin I B, Grishin N V, et al. Genome trees and the tree of life. Trends in Genetics, 2002, 18: 472~479[DOI]

12 Brown J R, Douady C J, Italia M J, et al. Universal trees based on large combined protein sequence data sets. Nature Genetics, 2001, 28: 281~285[DOI]

13 Qi J, Wang B, Hao B L. Whole genome prokaryote phylogeny without sequence alignment: a K-string composition approach. J Mol Evol, 2004, 58(1): 1~11[DOI]

14 Chu K H, Qi J, Yu Z G, et al. Origin and phylogeny of chloroplasts: a simple correlation analysis of complete genomes. Mol Biol Evol, 2004, 21(1): 200~206[DOI]

15 高　雷, 戚　继, 卫海滨, 等. 冠状病毒和人类 SARS 病毒的分子亲缘关系研究. 科学通报, 2003, 48(12):1170~1174

16 Matte-Tailliez O, Brochier C, Forterre P, et al. Archael phylogeny based on ribosomal protein. Mol Biol Evol, 2002, 19(5): 631~639

17 Doolittle R F, Handy J. Evolutionary anomalies among the aminoacyl-tRNA synthetases. Current Opinion in Genetic & Development, 1998, 8: 630~636[DOI]

18 Wolf Y I, Aravind L, Grishin N V, et al. Evolution of aminoacyl-tRNA synthetases-analysis of unique domain architectures and

phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Research, 1999, 9: 689~710

19 Woese C R, Olsen G J, Ibba M, et al. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiology and Molecular Biology Reviews, 2000, 64(1): 202~236[DOI]

20 Benson D A, Karsch-Mizrachi I, Lipman D J, et al. GenBank. Nucl Acids Res, 2003, 31: 23~27. Available at: ftp://ncbi.nlm.nih.gov/genbank/genomes/Bacteria/ [DOI]

21 Wheeler D L, Church D M, Federhen S, et al. Database resources of the National Center for Biotechnology. Nucl Acids Res, 2003, 31: 28~33. Available at: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/ [DOI]

22 Nei M, Kumar S. Molecular Evolution and Phylogenetics. New York: Oxford University Press, 2000. 87~103

23 Felsenstein J. PHYLIP (phylogeny inference package) version 3.5c, 1993, available at: http://evolution.genetics.washington.edu/phylip.html

24 Garrity G M, Johnson K L, Bell J A, et al. Taxonomic Outline of the Procaryotes. Bergey's Manual of Systematic Bacteriology. 2nd ed. Rel. 3.0. New York: Springer-Verlag, 2002. DOI: 10.1007/bergeysoutline200210 [DOI]

25 Burggraf S, Stetter K O, Rouviere P, et al. *Methanopyrus kandleri*: an archeal methanogen unrelated to all other known methanogens. Syst Appl Microbiol, 1991, 14: 346~351

26 Slesarev A I, Mezhevaya K V, Makarova K S, et al. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. Proc Natl Acad Sci USA, 2002, 99: 4644~4649[DOI]

27 Garrity G M, Bell J A, Lilburn T G. Taxonomic Outline of the Procaryotes. Bergey's Manual of Systematic Bacteriology, 2nd ed. Rel. 4.0. New York: Springer-Verlag, 2003. DOI: 10.1007/ bergeysoutline200310