

定向进化同源基因在细菌系统进化研究中的应用

蹇文婴 东秀珠*

(中国科学院微生物研究所 北京 100080)

关键词: 定向进化同源基因, 16Sr RNA, 细菌系统进化

中图分类号: Q93 **文献标识码:** A **文章编号:** 0253-2654(2000)05-0377-05

1 细菌系统进化及分类研究的发展——从表型到基因型

近年来,随着分子生物学理论和技术的发展,细菌系统进化及分类研究有了长足的进步。过去,判断某一细菌的分类地位,往往仅利用其表观特征,如细菌形态、菌落形态、生理特征、细胞组份等。随着分子生物学研究的深入,基因型特征成为细菌系统进化及分类研

究的主要依据。常用的基因型特征包括: DNA G + C mol%、DNA-DNA相关性分析、DNA 指纹分析及 rRNA 同源性分析等。值得一提的是, Woese 通过对各类生物 rRNA 序列进行分析,认为 ssu rRNA(16S 或 18S rRNA)序列是用于系统进化及分类研究最适宜的指标,

* 通讯作者

收稿日期: 1999-10-24, 修回日期: 2000-04-17

即分子计时器(Molecular chronometers)^[1]。通过对各类有机体的 ssu rRNA 序列进行比较, Woese 提出了真核生物(*Eukarya*)、细菌(*Bacteria*)和古菌(*Archaea*)的三域分类体系, 并构建了包括三域生命的进化树。目前, 16S rRNA 分析已成为生命系统进化及分类研究最常用的工具。

但是, 16S rRNA 序列分析也有其局限性。由于基因的水平转移及同一基因在不同进化分支中进化速度不同, 从单一分子序列来推测整个生物界的系统进化, 容易产生偏见^[2]。在 16S rRNA 系统发育树中, 某些部分的分支顺序不明确^[3]。而且, 单一分子序列所构建的生命进化树是没有根的, 这是因为, 寻找一群生物进化根需要比较该群生物与某一外群的关系。由于一个包括所有生物的生命进化树没有外群存在, 所以不可能用单一分子构建一个有根的生命进化树^[4]。另外, 16S rRNA 序列同源性分析比较适合于属以上的分类阶元间亲缘关系的研究, 而对于属以下的分类单位, 16S rRNA 序列分析的分辨率明显不足^[5]。

基于以上原因, 近年来, 一些进化中保守的生物大分子也被用于系统发育的研究。

2 用于系统进化研究的定向进化同源基因

正如生物进化学家曾经指出的那样, 生物大分子确实是生物进化历史的档案, 在真核生物中, 尤其是动物界, 生物大分子如血红蛋白、细胞色素 C 等早已被用来研究进化的历史。

近年来, 越来越多的生物大分子被用作研究系统发育、生命进化及分类学的工具。这些分子大部分是执行生命必需功能的酶类、辅酶或关键性的基因调控蛋白, 由于其功能保守, 它们的基因被称为管家基因(House-keeping gene)。从进化的角度来看, 这些基因又被称作定向进化同源基因(Orthologues)。所谓定向进化同源基因, 指的是从同一祖先垂直进化而来的基因, 它们在不同的物种中行使相同的、而且是生命活动所必需的功能。

一些已用于系统进化及分类学研究的定向进化同源基因列于表 1^[2]。

这些定向进化同源基因都具有作为分子计时器的基本特征^[1]: (1) 存在于三域或某一分支的所有生命体中, 可用来研究所有生物或某一分支的进化及生命的起源; (2) 功能保守, 进化缓慢, 变化速度可覆盖整个进

表1 用于系统进化及分类学研究的定向进化同源基因

| 功能 | 定向进化同源基因产物 |
|-------------|--|
| DNA复制和修复 | DNA旋转酶及拓扑异构酶、DNA聚合酶B、光修复酶、RecA蛋白 |
| 转录 | RNA聚合酶亚基、TATA结合蛋白、转录因子 |
| 翻译 | 延伸因子、氨酰tRNA合成酶、核糖体蛋白 |
| 中心代谢 | 3-磷酸甘油醛脱氢酶、3-磷酸甘油醛激酶、烯醇化酶、苹果酸脱氢酶 |
| 氨基酸合成及降解 | 转氨酶、谷氨酰氨合成酶、谷氨酸脱氢酶、精氨酸琥珀酸合成酶、氨甲酰磷酸合成酶、色氨酸生物合成酶类、组氨酸生物合成酶类、吡咯啉-5-羧酸还原酶 |
| 核酸合成 | 二氢叶酸还原酶、次黄苷酸脱氢酶、甲酰甘氨酸磷酸核糖合成酶 |
| 呼吸、光合及磷酸化作用 | 超氧化物歧化酶、过氧化氢酶、核酮糖-1, 5-二磷酸羧化酶、F ₁ F ₀ ATP合成酶、异化硫酸盐还原酶类、黄素蛋白、细胞色素 |
| 卟啉生物合成 | 谷氨酰氨tRNA还原酶 |
| 分子伴侣 | 热激蛋白 |
| 蛋白跨膜转运 | SecA、SecB |
| 细胞分裂 | FtsZ、微管蛋白 |
| 运动 | 鞭毛蛋白 |
| 信息传递 | 腺苷酸环化酶 |
| 其它 | 抗汞蛋白、DNA结合蛋白、磷脂酶C等 |

化历史, 且序列变化的速度与进化距离相当; (3) 分子中既含保守序列, 又含可变序列, 可用于研究进化程度不同的生物的系统发育; (4) 分子大小适中, 既能提供足够的统计学信息, 又适于研究中的操作。

系统进化研究中应用十分广泛的一类定向进化同源基因是热激蛋白(Heat shock protein)基因。热激蛋白广泛存在于细菌及真核生物细胞中, 是组成型表达的蛋白, 其主要功能是作为分子伴侣(Molecular chaperone)帮助蛋白质肽链折叠和组装成正确的构象, 这对细胞正常生长和增殖是必需的。当细胞受到理化刺激时, 它们的表达可进一步增强。根据其分子量, 热激蛋白可分为 HSP60、HSP70 等家族。HSP60 又叫伴侣蛋白(Chaperonin)、Cpn60、GroEL、60-65kD 抗原、细菌共同抗原(Common antigen)。分子伴侣在真核生物细胞质、线粒体及古菌中的类似物分别为 Tcp-1 (t-complex polypeptide), 线粒体 P1 蛋白和 TF55。由于分子伴侣是一类重要的抗原, 对它的研究成为近年来

的热点。许多生物分子伴侣基因序列已被测定, 这些生物包括革兰氏阳性细菌、衣原体、螺旋体、蓝细菌、紫细菌、植物及动物等。目前, GenBank、SwissProt 等数据库中收集了大量的分子伴侣序列, 为系统进化研究提供了丰富的素材, 利用分子伴侣进行的系统进化研究也已取得了很大进展^[6]。和分子伴侣一样被广泛应用于系统进化研究的定向进化同源基因还有 F_0F_1 ATP 合成酶、翻译延伸因子等。

同 16S rRNA 一样, 在利用定向进化同源基因进行的系统进化研究中, 通用的方法是比较核酸或氨基酸序列的同源性。但由于密码子的简并性, 基于氨基酸序列的同源性往往高于核酸序列的同源性。通过软件分析可以构建不同蛋白的系统进化树, 并推测进化的历史事件。另外, 特征序列 (Signature sequence 或 indel) 也被用来研究系统进化^[7]。特征序列指的是有一定长度的、保守的、存在于不同物种定向进化同源基因中相同位置上的缺失或插入序列。可以认为, 特征序列来自于进化过程中的某一个遗传事件, 并保留在发生了这个变化的祖先及其所有后代中。一般认为, 凡是含有该特征序列的生物来自于同一祖先, 而不含该特征序列的生物在该序列引入之前就分支出来或是来自于另一祖先。此外, 生物大分子的三维构象、免疫反应、指纹分析等也被用于系统进化及分类的研究。

3 定向进化同源基因所揭示的生物进化

生物大分子的分析已在生物系统进化及分类学研究的各个方面、各个层次得到了广泛的应用, 并取得了一定的进展。下面综述一些研究成果。

3.1 寻找生命的共同祖先 研究进化根本的问题是确定生命进化树的形状, 而确定生命树的形状, 就必须找到生命树的根, 即生命的共同祖先。前面已经提到, 由于没有外群, 由单一生物大分子构建的生命树是无根的。为了解决这个问题, Iwabe 等提出了利用平行进化同源基因 (Paralogues) 分析生物进化的方法^[8]。他认为, 虽然找不到所有生物的外群生物, 但可以找到一个外群基因, 该外群基因来自于 Cenancestor 即生命共同祖先中所发生的一次基因复制。可以设想, 一次古老的基因复制产生了基因 A 和 A', 它们保留在 Cenancestor 及其所有后代中。A 或 A' 可分别用来构建一个无根的包含所有生物进化树, 而 A 可作为 A' 的外群去寻找 A' 进化树的根, 反之亦然。Iwabe 选择了翻译延伸因子 EF-1 α / Tu 和 EF-2 / G 及 F_1F_0 ATPase 的 α 和 β 亚基这两对复制基因进行进化研究, 认为在进化过程中细菌首先分歧出来, 古菌和真核生物亲缘关系较近。之后, Brown 和 Doolittle 利用相同原理分析了数个氨酰 tRNA 合成酶的序列^[9], 得到相似结论, 并构建了以 Cenancestor 为根的进化树 (图 1.a), 在该树中古菌和真

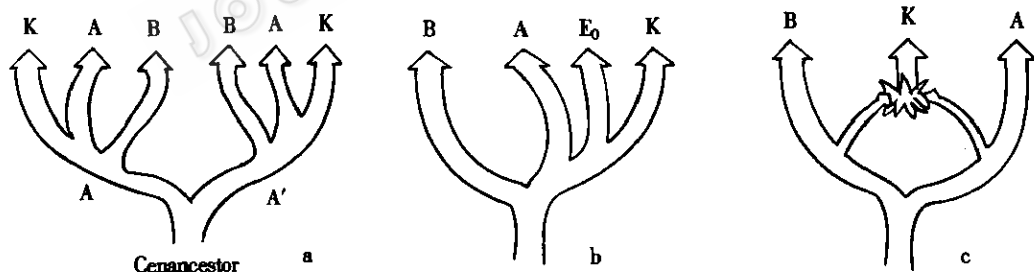


图1 根据不同生物大分子推测的系统发育树

a Brown和Doolittle的系统发育树, b Rivera和Lake的系统发育树, c Sogin和Zilg的系统发育树

(K 真核生物, A 古菌, B 细菌, E₀ 嗜泉古菌)

核生物为姐妹进化分支。

3.2 推测生命三域之间的关系 确定生命三域之间关系的中心问题是确定进化树的分支形状。

Rivera 和 Lake^[10]分析了核糖体的形状及 rRNA 序列, 并且根据真核生物和嗜泉古菌的 EF-1 α 中都含有一个 11 个氨基酸的特征序列而其它生物中不含该序列的现象, 提出了 Eocyte 生命树 (图 1.b), 在该树中 Eocyte

即嗜泉古菌 (*Crenarchaeota*) 与真核生物最近, 因而推测真核生物有可能起源于嗜泉古菌。

Sogin 和 Zilg^[12]认为真核生物并不是由古菌的祖先直接进化生成, 而是细菌或原核生物与古菌融合生成了真核生物的细胞核, 这就是真核生物的“嵌合起源学说” (图 1.c)。Gupta 和 Golding^[11]构建了 24 个蛋白的无根的生命进化树, 这些蛋白树支持嵌合起源学说, 他们

认为真核生物来自古菌和革兰氏阴性细菌的融合。

Brown 和 Doolittle^[2]综合分析了 66 个保守生物大分子的序列,构建了它们的系统发育树,这些树具有不同的拓扑形。其中,大部分参与 DNA 复制、转录、翻译的蛋白支持古菌和真核生物的姐妹进化关系;参与糖代谢的酶类则支持细菌和真核生物的姐妹进化关系;其它参与生物合成的蛋白及分子伴侣等则支持古菌与细菌亲缘关系较近。

Gupta^[7]认为,尽管 Brown 和 Doolittle 所分析的 66 个蛋白树揭示的结果并不明确,但大部分蛋白树支持古菌与革兰氏阳性细菌有较近的亲缘关系。Gupta 分析了存在于不同生物 HSP70、核糖体蛋白 S12 及 EF-1 α / Tu 等分子中的特征序列,发现古菌与革兰氏阳性细菌形成一个进化分支。根据以上事实,及革兰氏阳性细菌与古菌都只有单层细胞膜而革兰氏阴性细菌有双层细胞膜,提出了一种新的原核生物分类系统。在该系统中,所有的原核生物被分为单层细胞膜亚域 (Subdomain monodermata) 和双层细胞膜亚域 (Subdomain didermata),其中后者包括所有的革兰氏阴性细菌,而前者分为两个亚域 (Subsubdomain): 古菌 (Archaeobacteria) 和革兰氏阳性细菌 (Gram-positive bacteria)。古菌又分为广域古菌门 (Division Euryarchaeota) 和嗜泉古菌门 (Division Crenarchaeota); 革兰氏阳性细菌又分为低 G + C 门 (Division low G + C) 和高 G + C 门 (Division High G + C)。Gupta 认为这个分类系统既有基因型的依据 (蛋白序列分析),又有表型依据 (细胞膜层数),是一个较客观的分类系统。

3.3 推测光合作用的起源及进化 光合作用的出现是生命历史进程上的中心事件,也是进化研究中的一个难题。为了深入探讨这个问题,就必须确定具有光合作用的不同生物之间的进化关系。除了植物的叶绿体,几乎所有具有光合作用的生物都属于细菌,包括:螺旋菌、绿色非硫细菌、绿色硫细菌、蓝细菌、紫细菌。在 16S rRNA 系统树中,光合生物的分支分散在不具有光合作用的生物分支中,而且这一进化树中细菌分支的顺序并不十分明确。因而有必要选择其它生物大分子为工具,对光合生物进行研究。

Xiong 和 Inoue^[12]等对运动螺旋菌 (*Heliobacillus mobilis*) 的光合作用基因簇进行研究,确定了 30 个开放阅读框架,它们分别编码细胞色素、光反应中心蛋白等

光合作用组分。通过与其它光合生物的相应组分的序列进行比较,他们构建了多种蛋白的系统发育树。发现不产氧的光合生物 (包括只含 PSI 的紫细菌、绿色非硫细菌及只含 PSII 的弯曲菌、绿色硫细菌) 与产氧的光合生物 (同时含两种光反应系统的蓝细菌和绿色植物) 形成两个独特的分支,而且,前者更接近系统发育树的根部,是后者的祖先。其中,紫细菌是最早出现的光合作用分支。根据对系统发育树的分析,Xiong 及 Inoue 提出了光合作用的“异源融合 (Heterologous fusion) 起源学说”,即不产氧的光合生物的 PSI 与 PSII 某些组分融合,形成了产氧的光合生物的 PSII,该 PSII 再与不产氧的光合生物的 PSI 的某些组分融合,共同形成了产氧的光合生物的光合系统。

Gupta^[13]的观点却不相同,他选择 HSP60、HSP70、FtsZ 及丙氨酸氨 tRNA 合成酶研究光合作用的起源和进化,通过分析不同光合生物的这些蛋白中的特征序列,得出结论认为不同的光合生物由同一个祖先进化而来,其分支顺序是:螺旋菌→绿色非硫细菌→蓝细菌、叶绿体→绿色硫细菌→紫细菌。

3.4 支持内共生假说 真核生物的线粒体和叶绿体的内共生假说长久以来已得到了认可,该假说认为光合细菌进入发酵型的细菌细胞内,二者建立了一种永久性的共生关系;而且这种光合细菌又进而进化成叶绿体。线粒体则由好氧呼吸型细菌进入其它细菌细胞内进化而来。

16S rRNA 及细胞色素 C 的序列分析支持以上假说,但此假说的进一步验证,还需要其它依据。这方面,既存在于细胞器内又存在于细菌内的定向进化同源基因可提供有价值的进化信息。对 HSP60 和 HSP10 序列同源性分析表明,线粒体来自于紫细菌的 α 亚群的一个成员,而叶绿体来自于蓝细菌。另外 α 紫细菌和线粒体 HSP60 中含有同一种特征序列,而蓝细菌和叶绿体的 HSP60 含有另一种特征序列,这种现象也支持了细胞器的内共生假说^[6]。

3.5 作为 16S rRNA 分析的印证及补充,用于多种微生物系统发育及分类的研究 虽然目前研究微生物系统发育及分类的主要手段仍是 16S rRNA 序列的同源性分析,但已有越来越多的定向进化同源基因被用于系统发育研究,其结果多数印证了以 16S rRNA 为基础的结论,并揭示或阐明了 16S rRNA 分析无法解释的问题。

Viale^[3]等分析了来自于紫细菌、拟杆菌、衣原体、螺旋体、革兰氏阳性细菌、蓝细菌等 58 个细菌的 HSP60 蛋白序列,构建了一个无根的系统进化树,发现其聚类形式与 16S rRNA 树基本相似,从而印证了建立在 16S rRNA 同源性基础上的细菌分支的系统发育框架。此外,还阐明了一些 16S rRNA 分析中不明确的问题。例如,用最大似然量法(Maximum-likelihood method)构建的 16S rRNA 系统发育树中,高 G+C 和低 G+C 革兰氏阳性细菌具有共同的祖先,但 Bootstrap 评价分析却不能肯定这个结论。Viale 研究发现,与大肠杆菌的 HSP60 蛋白不同的是,所有革兰氏阳性细菌的 HSP60 中都有一个氨基酸的缺失,即特征序列,因而支持高 G+C 和低 G+C 革兰氏阳性细菌有共同祖先的结论。另外,还发现革兰氏阳性细菌与蓝细菌-叶绿体有很近的亲缘关系,这是 16S rRNA 同源性分析所没有揭示的。

Kwok^[14]等利用 HSP60 基因的部分核酸序列研究葡萄球菌属的系统发育。他们发现种内的序列相似性为 91%~98%,种间的序列相似性为 74%~93%(平均为 82%);而该属成员与枯草芽孢杆菌(*Bacillus subtilis*)及大肠杆菌(*Escherichia coli*)的最高相似性分别为 70% 及 59%。细菌分类学家普遍认为,16S rRNA 的同源性大于 97.5% 的细菌菌株可视为同种,而同属的 16S rRNA 序列同源性应大于或等于 90%。显然 16S rRNA 的高度保守性对于种内不同菌株间的鉴别是不适宜的,而在某些属内不同种间的系统发育研究中也存在分辨率不足的问题。对于种内株间或属内种间的亲缘关系研究,DNA-DNA 相关性分析能够提供足够的分辨率,但需要分析有关菌株或有关种两两之间的 DNA-DNA 相关性,是一项非常繁琐的工作。而 HSP60 的分析既能提供较高的分辨率,又便于操作,有可能成为研究种内或种间亲缘关系适宜的工具。

Kamla^[15]等利用延伸因子 EF-Tu 的核酸序列构建了支原体的系统发育树,发现分支的划分及相互关系与其成员的表型特征(细胞组分、代谢途径等)十分吻合,而 16S rRNA 的系统发育树却不具有这样的相关性,说明 EF-Tu 树比 16S rRNA 树更准确的反映了支原体各成员表型和基因型之间的联系,在此属的种间亲缘关系研究上是一种更客观的分类标准。

除以上几例之外,还有利用鞭毛蛋白研究疏螺旋

体(*Borrelia*)的系统发育;利用 RecA 蛋白研究绿菌属(*Chlorobium*)的系统发育;利用磷脂酶 C 研究梭菌属(*Clostridium*)的系统发育,等等。它们已成为 16S rRNA 同源性分析的印证及补充,使系统发育研究更加客观、准确。

4 生物大分子序列分析的前景与展望

从前面的讨论可以看出,生命的共同祖先是什么?三域生命形式的进化关系是什么?以及各个进化分支内的进化顺序是什么?等等,这些问题仍处于推测之中。由于基因的水平转移及各个进化分支进化速度的差异,导致了不同蛋白质所构建的系统发育树不同,它们支持不同的结论。从统计学的角度来看,由大部分的蛋白系统发育树所支持的结论有可能是更准确的。所以,生命进化问题的最终揭示有赖于寻找出更多的、适合于作为分子时钟的蛋白,构建更多的系统发育树。

自 1995 年流感嗜血杆菌(*Haemophilus influenzae*)和生殖道支原体(*Mycoplasma genitalium*)全基因组序列发表以来,系统进化的研究就进入了比较基因组时代。1996 年第一个古菌——詹氏甲烷球菌(*Methanococcus jannaschii*)的基因组序列的完成使得利用基因组分析来研究生命三域的进化成为可能。通过寻找全基因组序列中的开放阅读框架,预测蛋白质功能,并比较不同生物定向进化同源基因的开放阅读框架及不同开放阅读框架的排列顺序,可以全面而客观的判断这些生物的系统进化关系。相信越来越多的生物的全基因组序列终将会使人类找到生命进化的根源,揭开生命之谜的谜底。

参 考 文 献

- [1] Woese C R. Microbiol Rev, 1987, 51:221~271.
- [2] Brown J R, Doolittle W F. Microbiol Mol Biol Rev, 1997, 61:456~502.
- [3] Varian A M, Arakaki A K, Soncini F C, et al. Int J Sys Bacteriol, 1994, 44:527~533.
- [4] Doolittle W F, Brown J R. Proc Natl Acad Sci USA, 1994, 91:6721~6728.
- [5] Stackebrandt E, Goebel B M. Int J Sys Bacteriol, 44: 846~849.
- [6] Gupta R S. Mol Microbiol, 1995, 15:1~11.
- [7] Gupta R S. Mol Microbiol, 1998, 29:695~780.
- [8] Iwabe N, Kuma K, Hasegawa M, et al. Proc Natl

(下转第 388 页)

(上接第 381 页)

- Acad Sci USA, 1989, **86**:9355~9359.
- [9] Brown J R, Doolittle W F. Proc Natl Acad Sci USA, 1995, **92**:2441~2445.
- [10] Rivera M C, Lake J A. Science, 1992, **257**:74~76.
- [11] Golding G B, Gupta R S. Mol Biol Evol, 1995, **12**: 1~6.
- [12] Xiong J, Inoue K, Bauer C E. Proc Natl Acad Sci

USA, 1998, **95**:14851~14856.

- [13] Gupta R S, Mukhtar T, Singh B. Mol Microbiol, 1999, **32**:893~906.
- [14] Kwok A Y C, Su S, Reynolds R P *et al.* Int J Sys Bacteriol, 1999, **49**:1181~1192.
- [15] Kamla V, Henrich B, Hadding U. Gene, 1996, **171**: 83~87.