

文章编号: 0529-6579 (1999) 01-0001-06

生物分子分类检索表——原理与方法^{*}

屈良鹄, 陈月琴

(中山大学生物工程研究中心, 广州 510275)

摘要: 在对大量分子数据比较分析的基础上, 对分子检索表这一概念及其遗传学原理进行了探讨, 并根据 rRNA 基因结构特点及进化规律讨论了生物高级阶元以及生物物种界定的分子标准. 分子检索原理和方法的阐明, 不仅对生物资源的分类鉴定有重要的价值, 而且在医学, 生态学, 古生物学等许多领域中有广泛的应用前景.

关键词: 分子检索表, rRNA 基因, 分子特征

分类号: Q 19 Q 751 **文献标识码:** A

传统的生物分类系统主要是根据生物的形态构造和生理生化特点为基础的自然分类系统. 即根据生物所具有的特征, 将生物逐级分类, 建立了一系列分类阶元. 各级分类阶元的建立, 不仅构成了一个完整的, 相互联系的生物系统; 同时也为各种生物鉴定规定了一系列的标准和方法, 即生物检索表. 这些标准和方法已成为人们认识自然界的有力工具. 随着对生命本质探索的深入以及新的分子生物学方法和技术的突破, 分子数据大量积累并与日俱增, 使人们完全能够从分子水平来认识生物物种分化的内在原因和物质基础以及各类生物的分子进化历史, 从而导致了新的分子系统的建立. 其意义已超出了对传统生物系统作一个简单或局部的修正, 而且有可能整个地改变传统生物学的许多基本概念和方法^[1].

那么, 根据分子数据重建的系统与传统分类系统关系如何? 根据分子特征是否也可以建立起相应的鉴定与分类标准, 即生物分子分类检索表 (简称分子检索表)? 本文在对大量分子数据比较分析的基础上, 对分子检索表这一概念及其遗传学原理进行了探讨, 并根据核糖体 RNA (rRNA) 基因结构特点及进化规律讨论了生物高级阶元以及生物物种界定的分子标准.

1 概念提出的遗传学原理及检索方法

核酸分子 (DNA 和 RNA) 包含着生物遗传的信息, 同时也记载着生物进化的历史, 即核酸结构不仅指导生物的形态建成和个体发育, 而且还能够反映生物之间的亲缘关系.

^{*} 基金项目: 国家杰出青年基金 (39525007); 国家自然科学基金 (39770058)

收稿日期: 1998-10-15 第一作者简介: 屈良鹄, 男, 45 岁, 教授

目前生物分子系统的构建，都是基于这一原理而进行的²¹。

比较生物分子分类系统与传统的生物分类系统，尽管所依据的分类特征完全不同，但是两种系统在许多方面都具有极大吻合性。由此可见传统分类学所强调的分类特征必然存在着其基因水平的分子基础。但是，与传统分类学不同，分子系统是以 DNA 序列作为唯一类比性状，而传统分类学则是根据多元性状的综合分析来进行的。根据分子进化的规律，任何一种生物的基因组中，必然存在该生物物种独有的特征性核苷酸序列，同时也包含着该生物所属类群的共有核苷酸序列（即共征）。分子检索表就是由反映各种生物分类特征的 DNA 序列所构成。对不同生物 DNA 序列的比较，相当于对这些分类特征进行了一次综合的分析及检索。这就是分子检索表概念的遗传学原理。

目前分子检索的方法主要是通过比较生物基因组或某一具有代表性基因的序列总体相似性，或基因的特征核苷酸序列及高级结构，对某一种生物进行鉴定，并确定其分类学位置。由于大量已知生物的基因序列都已被测定，并输入到国际基因数据库。如基因数据库中现已收集 280 万条核酸序列（Genbank，1998 年 4 月）。利用国际基因数据库或建立专门的分子分类数据库，可以作为各种生物分子检索的参照系统。只要测定一个未知生物样品的 DNA（或 RNA）序列，并通过在这一参照系统中的检索，就可以对其进行分类鉴定。因此，采用核酸序列来作为分子指标进行分子检索研究，不仅能够为生物的分类鉴定建立起一个简单而统一的标准；而且将大大促进对生物物种分化机制的认识以及分子分类高级阶元的确立。

2 通用的分子检索指标——大分子 rRNA 基因

直接地比较各种生物基因组 DNA 序列，将能够提供最具有说服力的分子检索结果，但却是一项浩大的工程。目前，采用基因组的一个片段（包含一个或几个基因的序列）来作为分子指标，是可行的，因为在某种意义上来说，认识一棵树不必摘下所有的叶子。许多基因如脊椎动物中的血红蛋白基因，植物叶绿体 1，5 二磷酸核酮糖羧化酶大亚基（Rb-cL）基因等都已成功地用于生物的分类与系统学研究。但是由于它们仅存在于某些生物类群中，在应用上有一定的局限性。相对来说，大分子 rRNA 基因（编码 16S 类 rRNA 和 23S 类 rRNA），由于其普遍存在于所有的生物细胞中，并且具有相似的结构和功能，是目前可

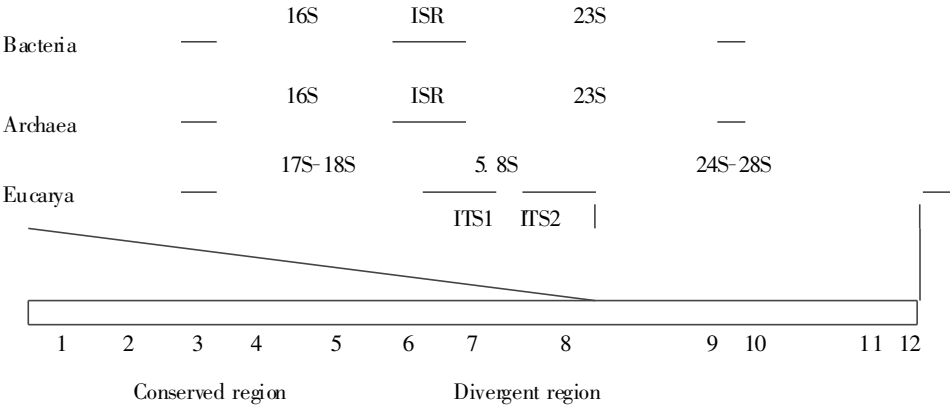


图 1 大分子 rRNA 基因结构模式图

Fig. 1 Diagram of rRNA gene structure

用于所有生物比较分析的通用的分子检索指标.

		3 400			3 439
Eucarya	<i>Mus musculus</i>	TAAACGGCGG	GAGTAACTAT	GACTCTCTTA	AGGTAGCCAA
	<i>Oryza sativa</i>	-----	-----	-----	-----
	<i>Saccharomyces cerevisiae</i>	-----	-----	-----	-----
	<i>Physarum polycephalum</i>	-----	-----	-----	-----
Archaea	<i>Methanococcus vannielii</i>	-----	-G-----	A--C-----	-----G--
	<i>Halobacterium halobium</i>	-T-----	-G-----	---C-----	-----GT-
	<i>Thermococcus celer</i>	-----	-G-----	A--C-----	-----G--
	<i>Thermoplasma acidophilum</i>	-----	-G-----	A--C-----	-----G--
Bacteria	<i>Escherichia coli</i>	-----	CC-----	A--GG--C--	-----G--
	<i>Bacillus stearothermophilus</i>	-----	CC-----	A--GG--C--	-----G--
	<i>Ralstonia pickettii</i>	-----	CC-----	A--GG--C--	-----G--
	<i>Zoogloea ramigera</i>	-----	CC-----	A--GG--C--	-----G--

图 2 rRNA 基因保守区中的特征序列与生物高级阶元的分子标准

Fig 2 Characteristic sequences in the conserved region of rRNA gene used for the molecular criteria of higher hierarchy delimitation

另外, 大分子 rRNA 基因的结构特点又决定了它是一个良好的指标. 图 1 示大分子 rRNA 基因结构模式图. 从中可看出大分子 rRNA 基因中存在着进化速率不同的结构区, 主要可分成两类即保守区和高变区. 保守区是指在亲缘关系较远的生物类群 (如动植物) 之间比较, 其大分子 rRNA 基因某一段序列中发生的核苷酸变异较少 (例如序列相似性在 70% 以上). 高变区恰恰相反, 即使在两个亲缘关系十分接近的物种之间比较, 在这些区域中也存在核苷酸差异; 而且, 这些差异会随着比较物种的亲缘关系的疏远而迅速扩大. 并伴随着核苷酸的插入或缺失, 所以在不同生物类群之间, 大分子 rRNA 基因高变区的序列差别很大, 甚至完全不同. rRNA 基因中的保守区和高变区是交替分布的. 大分子 rRNA 基因这一进化模式, 在用于亲缘关系远近不同的生物特征分析时具有极大的优越性^[3].

目前, 国际基因数据库中已收集 50 591 条各种生物的 rRNA 基因序列. 由于其增长速度还在加快, 可以预见, 在不远的将来, 国际基因数据库将包括地球上几乎所有细胞生物的 rRNA 基因序列, 使之成为一个对所有细胞生物进行鉴定和分类的参照系统. 由于 rRNA 基因分析的快速和准确性, 所以这一方法已成为现今主要的分子检索手段.

3 rRNA 基因中的特征序列与生物高级阶元的分子标准

根据 rRNA 基因结构进化模式, 在 rRNA 基因保守区中, 有可能存在着与分类高级阶元有一定对应关系的核苷酸序列. 如在 23S 类 rRNA 基因中, 我们发现 C4, C5, C9 和 C11 为 4 个高度保守区^[4], 从中可以鉴别出一系列存在于所有细胞生物中的共有序列, 如 5' GAGCTGGGTTYARAMCGTYGTGAGACGAG 3'. 这些共有序列, 不仅表明所有细胞生物起源于共同的祖先, 同时也是地球上细胞生物界共同的分子特征. 在高度保守区 C9 中, 还可以鉴定出真核生物特征性序列 (图 2), 从图中可以看到, 长达 40 个核苷酸的序列在所分析的真核生物 (从人到粘菌) 中完全相同, 而与两类原核生物 (细菌和古细菌) 的序列却存在明显的差异, 因此可以作为真核生物界的分子特征. 在该高度保守区中, 也很容易定义细菌和古细菌界各自的特征性序列, 而在变化区中, 则有可能存在着较低阶元的分子特

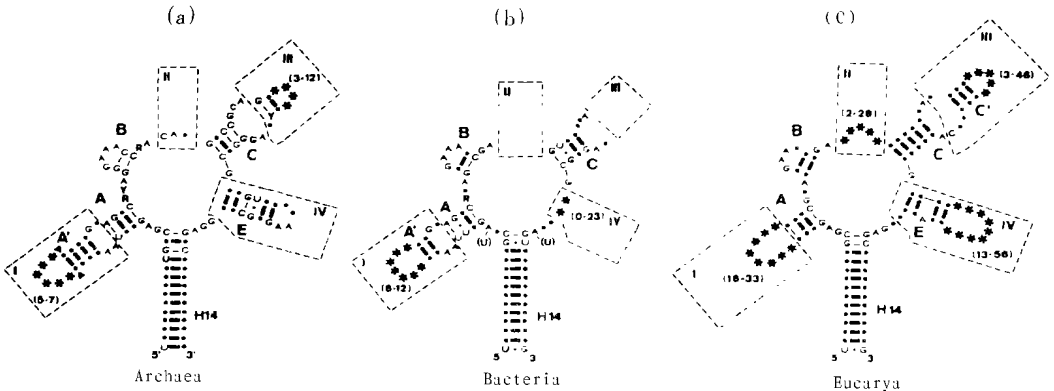


图 3 D3 变化区的二级结构模式

Fig 3 Secondary structure of 23S-like rRNA D3 region

征。例如，虽然在 23S 类 rRNA D2 变化区中，2 种老鼠（*Mus mulculus* 和 *Rattus rattus*）的序列表现出极大的差异，但是在 23S 类 rRNA D1 变化区却有一段长达 100 个核苷酸在哺乳动物中保守不变的特征序列。最近我们从细菌的比较中，发现仅存在于各种蓝藻 16S—23S rRNA 基因转录间隔区（ISR）中一个非常保守的 20 个核苷酸的序列 5' GAACCTTGAAAAC-TGCATAG 3'，这一序列很可能具有重要功能意义，同时也是蓝藻的一个分子特征（待发表）。通过对沙门氏菌与大肠杆菌 16S—23S rRNA 基因转录间隔区的比较分析，一段长达 70 个核苷酸的沙门氏菌属特征性序列已被鉴定出来^[5]。因此，选择各种生物 rRNA 基因的不同区域进行比较分析，就有可能发现与某种生物类群相关的特征性核苷酸序列或突变。显然，这些特征性核苷酸序列可以作为检验其相应的分类阶元的分子标准。

分子特征不仅指特征性序列的鉴定，rRNA 高级结构也能反映出生物类群的特点，从而具有分类学价值，例如大分子 rRNA 所有的保守区可以形成一个在各种生物中都相同的核心结构，但是在变化区中却可以鉴别出反映不同生物特征的二级结构。如在 23S 类 rRNA 的 D3 变化区中就可以鉴别出反映真核生物，细菌和古细菌特征的二级结构（图 3）^[9]。

除了特征性序列之外，RNA 基因中一个结构区域的长度特点，也能反映某一类生物在进化中的共同特征，例如，我们发现裸子植物 18S—25S rRNA 基因转录间隔区（ITS）长度在 1 100 碱基对以上，而苔藓、蕨类、被子植物的 ITS 片段长度却都在 800 碱基对以下，因此，ITS 长度特异性，可以作为检验裸子植物门的一个分子标准（图 4）^[7]。

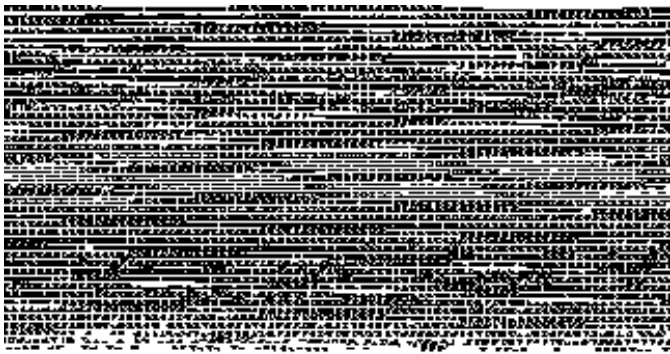


图 4 裸子植物 rRNA 基因 ITS 区长度特异性

Fig 4 A charactenistic size of the rRNA gene ITS in Gymnosperm

M 2 kb marker; 1 买麻藤; 2 苏铁; 3 银杏;
4 水松; 5 木麻黄; 6 毛茛; 7 灰木莲; 8 假鹰爪;
9 毛茛泽泻; 10 华南毛蕨; 11 长蒴丝瓜苔藓

rRNA 基因中的特征性序列和结构的发现, 为一些高级分类阶元的分子标准的建立提供了依据. 但是, 分子系统与传统分类系统在高级阶元上并不可能完全对等, 这是因为宏观的形态进化与微观的分子进化并不一定同步. 同时在传统分类高级阶元的设立上, 也有一定的主观性. 对各种生物 rRNA 基因和其他基因的进一步研究, 将使我们生物高级阶元的分子标准有更深入的认识.

4 物种界定的分子指标

物种是分类的基本单元. 物种的划分, 主要是根据形态和生殖的标准. 对应这些标准的分子基础是什么? 这是一个十分复杂的问题, 因为造成形态变异和生殖隔离的因素往往是多种多样的. 但是通过对 rRNA 基因序列的种间异质性和种内均一性的分析, 我们认为 rRNA 基因的转录间隔区 (ITS 区) 可以作为物种界定的一个分子指标. rRNA 基因是一类多拷贝基因, 由于细胞中存在一种协同进化机制^[8], 即借助于染色体的不平等交换, 基因的转换和转位等分子校正机制, 并在选择压力作用下, 逐步将细胞中各个 rRNA 基因拷贝中发生的突变固定并分布到所有的 rDNA 拷贝中去, 从而形成了同种生物 rRNA 基因各个拷贝的均一性. 这一进化机制为采用 rRNA 基因 ITS 区作为物种界定的指标提供了基础. 同时, 由于 rRNA 基因 ITS 区的异质性, 使得其在各类生物属种间表现出明显的差异, 这也是目前国际上采用 ITS 区作为各种生物类群属种间分类与系统学研究的主要原因.

然而, 由于 ITS 区在各类生物物种间的变化程度不同, 所以, 目前仍不能以一个统一的数值作为物种界定的标准, 但是可以根据不同生物类群的特点来制定 ITS 序列差异百分比值, 作为某类生物种间划分的一个分子标准. 我们通过 Internet 从国际分子生物学数据库中获取不同生物类群的 rRNA 基因 ITS 序列, 比较得出: 被子植物大多数科属其 ITS 序列的种间差异值为 1.2%~10.2%, 属间差异值为 9.6%~28.8%; 真菌和藻类中, 种间差异值一般大于 14.0%; 某些单细胞微藻类, 种间差异值可达 20%以上. 随着进一步的研究, 一个物种界定的分子标准将会更加完善.

5 结 语

分子检索标准与方法的提出和原理的阐明, 从分子水平为地球各级类群生物分类鉴定提供了客观依据. 这一方法的建立是分子生物学与分类学相互渗透的结果, 并将有助于人们对自然界更加深入的认识. 尤其是以 rRNA 基因特征核苷酸序列为基础的生物系统探针——rRNA 探针群, 将成为一种新的生物鉴定和分类方法^[9], 它们可被用于特异性检测某一种或某一类群生物的存在, 并同时鉴定它们所属的系统发育类群. 这不仅对生物资源的分类鉴定有重要的价值, 而且在医学、生态学、古生物学等许多领域中也有广泛的应用前景.

参考文献:

- [1] WOESE C R, KANDLER O, WHEELIS M. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci (USA)*, 1990, 87: 4576~4579

- [2] ZUCKERKANDL E, PAULING L. Evolutionary divergence and convergence in proteins. In: BRYSON V, et al. (C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

- al eds. *Evolving genes and proteins*. New York and London: Academic Press, 1965. 97~166
- [3] QU L H, NICOLSO M, BACHELLERIE J P. Phylogenetic calibration of the 5' terminal domain of large rRNA achieved by detemining twenty eucaryotic sequences. *J Mol Evol*, 1988, 28: 113~124
- [4] QU L H. Structuration et evolution de L' ARN ribosomique 28S chez les eucaryotes. [学位论文]. Toulouse: These de Doctoral d'Etat Universite Paul-sabatier. No. 1273, 1986
- [5] 周惠, 屈良鹄, 陈月琴, 等. 肠炎沙门氏菌两种 rDNA 16S-23S 基因间隔区序列分析. *中山大学学报 (自然科学版)*, 1997, 36 (5): 74~77
- [6] MICHOT B, QU L H, BACHELLERIE J P. Evolution of large subunit rRNA structure. The diversification of divergent D3 domain among major phylogenetic groups. *Eur J Biochem*, 1990, 188: 219~229
- [7] QU L H, XI B Q, SHI S H, et al. A characteristic size of the ITS in gymnosperm rDNA-revealed by a rapid PCR method. Tokyo: X V Intemational Botanical Congress. 1993. 201
- [8] GERBI S. The evolution of eukaryotic ribosomal DNA. *Biosystems*, 1986, 19: 247~258
- [9] 屈良鹄. 微生物系统发育的分子遗传学研究概述. 见: *微生物遗传学研究综述集*. 上海: 复旦大学出版社, 1993

Key to Molecular Taxonomy —— Principles and Methods

QU Lianghu^{*}, *CHEN Yueqin*

Abstract: Based on the analyses of rapid increasing molecular data, the principles and methods of molecular criteria for taxonomy were approached. And according to the structural characteristics and evolutionary modes of rRNA gene, the molecular criteria for higher hierarchy and species delimitation were suggested, which not only have a great value in the identification and classification of biological resources, but also show a perspective in the fields of medicine, ecology and archaeology.

Keywords: molecular criteria, rRNA gene, taxonomy

* Biotechnology Research Center, Zhongshan University, Guangzhou 510275, China