

# 16S rRNA 二级结构的研究进展及其在系统分类中的应用

陈国忠, 李文均, 徐丽华, 姜成林

(云南大学 云南省微生物研究所 教育部微生物资源开放研究重点实验室, 云南 昆明 650091)

**摘要** 16S rRNA 作为研究系统进化的生物大分子极受重视, 是系统分类中必不可少的一个指标。然而, 在进行系统发育分析时通常的序列联配方法有明显的局限性, 即无法摆脱一级结构的束缚。20世纪80年代以来, 一些研究者把目光投向了16S rRNA 二级结构的分析研究, 并且已经取得了相当不错的进展。研究表明, 16S rRNA 二级结构可能作为一个新的分类指标, 成为系统分类的一个重要补充。

**关键词** 16S rRNA; 二级结构; 系统分类

中图分类号: Q939.13 文献标识码: A 文章编号: 1005-7021(2005)05-0054-04

## Advanced in 16S rRNA Secondary Structure and Its Application in Systematic Classification

CHEN Guo-zhong, LI Wen-jun, XU Li-hua, JIANG Cheng-lin

(Key Lab for Microbial Resources under the Ministry of Education, Yunnan Inst. of Microbiol. Yunnan Univ. Kunming 650091)

**Abstract** 16S rRNA as macromolecule in studying organism systematic evolution has attracted close attention. It is an absolutely necessary index in systematic classification. However, when carrying out systematic analysis the general multi-sequence alignment method obviously has its limitation, i. e. it cannot shake off its primary structure. Since 1980s, some researchers have thrown their look at 16S rRNA secondary structure analysis study, and have made a great progress. The study suggested that 16S rRNA secondary structure may become an important replenishment in systematic classification as a new classification index.

**Key words** 16S rRNA; secondary structure; systematic classification

在很多情况下, 对于 RNA 来说其在结构构成上的生物学意义要大于它的序列组成, 这对于非编码 RNA 或结构 RNA 来说尤其如此<sup>[1]</sup>。很多同源的 RNA 有着相同或相似的二级结构或三级结构, 然而却在一级结构上很少有什么有意义的相似序列, 比较典型的的就是 16S rRNA。只要维持其原来的碱基互补状况, 即使对序列进行很大程度的补偿突变 (compensatory mutations) 常常也是可以容忍的, 即对于 16S rRNA 在功能上没有什么太大影响。也就是说, 16S rRNA 的同源性更多的是建立在其特定的二级结构具有保守性, 而不是具有哪一段保守的序列, 它所具有的功能活性也是由特定的结构所决定的。此外, 研究者在分

析了大量统计数据后, 发现 16S rRNA 二级结构具有重要的分类学意义, 它对于系统分类学的研究和发展意义重大。

## 1 16S rRNA 二级结构的研究进展

### 1.1 16S rRNA 二级结构模型的提出和发展

1981年, Woese 等<sup>[2]</sup>以 *E. coli* 的 16S rRNA 为研究对象, 综合化学修饰、蛋白结合、酶切片断等实验数据, 初步推断出了 16S rRNA 二级结构模型, 并对该模型的特征进行了简要描述。该模型的建立有助于阐明 16S rRNA 结构与功能的关系以及 30S 小亚基的组装方式, 揭示了在众多生物类群中核糖体不仅核苷酸序列保守, 结构也同

收稿日期: 2005-01-10

作者简介: 陈国忠 男, 在读硕士研究生。主要从事高温放线菌的分离及系统分类学研究。

基金项目: 国家“973”项目(2004CB719601); 国家自然科学基金(30270004)

样保守,并且在核糖体功能行使过程中结构的稳定性至关重要。这篇文章发表在 Science 上,引起了相关研究领域的极大关注。从此,越来越多的研究者投入到 16S rRNA 二级结构的研究中来,无论功能结合位点还是突变进化机制,无一不是研究热点。

Woese 认为,早期对于 rRNA 二级结构的研究都只限于部分序列,结果并不可靠(后来也证明是错误的)。1983 年, Woese 等<sup>[3]</sup>在原有的基础上,比较分析了真细菌域、古细菌域和真核生物域 150 多个物种的 16S rRNA 全序列,详细论述了三域生物 16S rRNA 二级结构的特点,发现序列比较方法不但是预测假设结构的一个有力工具,而且有利于发现该分子的重要功能区段;同时指出二级结构上的改变可能导致能量变化并影响到功能。此文论述如此详尽而准确,以至于近 20 a 来,涉及到 16S rRNA 二级结构的大多数文章仍将其作为重要参考文献之一。

随后,很多研究者也加入到 16S rRNA 二级结构的研究中来,对其模型不断改进和完善。比如 Neefs<sup>[4]</sup>在 1993 年发表的文章引用率也很高,文中简明扼要的描述了真细菌、古细菌和真核生物的核糖体小亚基的二级结构模式图以及相互之间的主要区别,这对于以后的理论研究具有重要的参考价值和指导意义。针对 16S rRNA 二级结构的模型,虽然在某些细节上意见有所分歧,但是至今仍以 Woese 等构建的二级结构模型为主。

## 1.2 预测方法

一般来说,通过 X 射线晶体衍射和 NMR 等实验方法测得的 RNA 分子的立体结构是最精确可靠的,但是由于 RNA 分子具有降解速度快、难以结晶等特点,这两种方法费时费力且代价高昂,故而借助于计算机手段和各种数学方法从理论上预测 RNA 的空间结构,是提高认识 RNA 空间结构效率的一个捷径,也是目前主要依靠的方法。

RNA 二级结构的预测主要包括 2 种方法:最小自由能算法和比较序列分析方法。最小自由能算法是目前最流行的一种算法, RNAstructure 软件是该算法的代表之作。自 1981 年由 Zuker<sup>[5]</sup>提出并发展至今,最小自由能算法已经相当成熟,使用起来非常方便,预测结果也是比较成功的,尤其是对于小分子的 RNA。可惜,该算法有个最大

的缺陷就是不能预测假结,而且目前还缺乏对多分支环和假结自由能参数的细致了解。尽管如此,最小自由能算法还是获得了广泛应用,根据其原理设计出的软件程序颇受相关领域研究者的青睐。

比较序列分析<sup>[6~8]</sup>通过多序列联配并结合各类统计分析和序列上下文语义分析,针对一系列相关序列构建出它们的一个通用二级结构模型,以次来推断未知序列的二级结构。作为一个理论结构预测方法,比较序列分析是最值得信赖的一种方法,其预测结果仅次于实验上用 X 射线或 NMR 测定的结构。但由于它是一种基于已有序列的先验知识的方法,因此首先要拥有一定数量的相关序列样本,并且假定这些序列应该具有一致的二级结构和一些共同的基本结构单元。对于小样本或者来源差异很大的序列,其比较结果就不大可靠了。

预测方法多种多样,总的一个趋势就是互相之间取长补短。通常采取的办法是把最小自由能算法和比较序列方法结合起来<sup>[9]</sup>,一般先比较后折叠,利用比较后的统计信息和先验知识,把与二级结构模型对应的同源序列剪切下来,再把它加入到折叠过程中来提高预测的准确性。

## 1.3 数据库

时至今日,16S rRNA 二级结构的相关数据库非常多,这里主要介绍其中一个:Comparative RNA Web Site<sup>[10]</sup>(<http://www.rna.icmb.utexas.edu/>)。该数据库是最权威、最专业的 RNA 数据库之一,相比其他一些类似的数据库,它内容丰富,数据全面而精确,且更新速度快。该数据库通过比较 RNA 序列分析方法,揭示 RNA 的高级结构和保守性,并构建系统树;提供 3 种核糖体 RNA (rRNA、mRNA 和 tRNA) 以及其他一些 RNA 的 4 类信息:最新的比较结构模型、核苷酸保守性、序列和结构的相关数据以及获得数据的方法系统。

16S rRNA 二级结构的数据信息是一个重要组成部分,主要包括: Noller-Woese-Gutell 16S rRNA 比较结构模型的进化分析,并通过将不同版本的 16S rRNA 二级结构中的碱基对列表,记录比较结构模型的改进;16S rRNA 碱基对和单个碱基的保守性分析;以及 16S rRNA 二级结构

相关的精美图表,并提供 PostScript,PDF 等多种格式下载。

## 2 16S rRNA 二级结构在系统分类中的应用

近年来,用于修正种属甚至更高分类级别的最有力工具就是 16S rRNA 序列及其二级结构分析,尤其是二级结构,可谓异军突起。相比一级结构,二级结构在系统分类中有着独特的意义。首先,16S rRNA 的二级结构比一级结构更为保守,甚至在古细菌和真核生物之间区别也很小,因此,二级结构的不同可能暗示着系统发育的多样性<sup>[11]</sup>;其次,16S rRNA 二级结构具有很多一级结构无法体现出来的特征(图 1),这些特征具有特殊而重要的生物学意义,并且表现出种属特异性,因此被越来越多的用于系统分类研究。

1992 年, Wisotzkey 等<sup>[12]</sup>对芽孢杆菌属(*Bacillus*)的 16S rRNA 序列进行系统发育分析,比较了 16S rRNA 部分同源区段的二级结构,同时结合脂肪酸类型,从中划分出一个新属:脂环酸芽孢杆菌属(*Alicyclobacillus*)。该属包含原 *Bacillus* 属 3 个种,它们除了具有嗜酸嗜热的特点和都以  $\alpha$ -alicyclic fatty acid 作为主要的脂肪酸组分之外,16S rRNA 几个区段的二级结构明显区别于芽孢杆菌属其它种,而三者之间则相同或者极为相似。例如:65~104 区段(参照 *E. coli* 的 16S rRNA<sup>[4]</sup>)的区别最为典型,最明显的变化就是少了一段长达十几个碱基的茎区(图 2)。

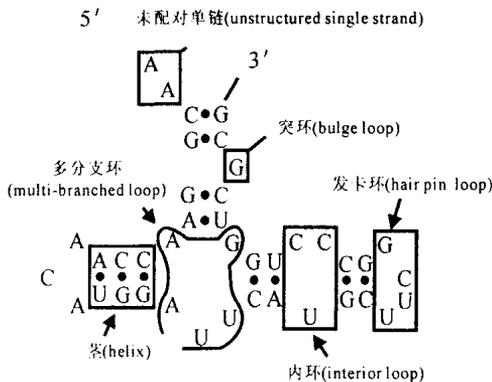


图 1 RNA 二级结构模型及其各结构单元名称

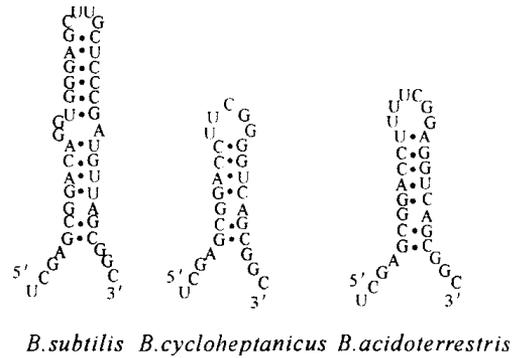


图 2 *B. subtilis* 为 *Bacillus* 属模式种,另外 2 个种原来在 *Bacillus* 属,后被归入 *Alicyclobacillus* 属

另外还有几个区段也有所不同,不过差别较小,比如 829~857 区,997~1 044 区和 1 128~1 144 区。这些二级结构上的特点使得三者亲缘关系显得更近而与 *Bacillus* 属其他种更远。

相比之下,国内这方面的工作起步较晚,且多放在预测方法方面,很少有人将 16S rRNA 二级结构运用于系统分类学研究。2003 年,陈朝银等<sup>[13]</sup>对 *Thermaceae* 科 38 株菌的 16S rRNA 序列进行了系统分析,重点研究了 16S rRNA 的特征性核苷酸和二级结构特征,将 38 株菌划分为 18 个种,此结果与 NCBI 中的分类基本一致(<http://www.ncbi.nlm.nih.gov/>)。郭春雷等<sup>[14]</sup>对所分离得到的几株栖热菌属(*Thermus*)的菌株和该属一些有效发表种进行了基于 16S rRNA 的系统发育分析,预测并比较了 16S rRNA 部分区段的二级结构,利用内环、内环与发卡环之间碱基对数目以及发卡环的碱基数目区分种间差异,证明 16S rRNA 二级结构可以用来区分种一级的分类单位。

即使不直接用于系统分类,16S rRNA 的二级结构也经常被用作序列联配时的参照。仅 2002 至 2004 三年中,在 IJSEM 就有数十篇文章提到在对 16S rRNA 序列进行联配时应参照二级结构<sup>[15~17]</sup>,即在多序列联配之后,根据 16S rRNA 二级结构对结果进行手动修改,以使同源序列较好的对齐,研究者认为这样构建出的系统进化树更为可靠。

### 3 结 语

一直以来,国内外很多研究者致力于二级结构的研究,不断有新的方法被研究出来用于 RNA 二级结构的预测<sup>[18,19]</sup>,虽然一些区段的结构仍然不确定,但是随着工作的深入和细致,16S rRNA 二级结构模型处于不断的完善之中,而随之带来的就是系统分类学的发展。现今,分类指标除了经典、化学、分子之外,还有酶切图谱、指纹图谱、随机扩增片断等,可谓百家争言,不过都有一定的缺陷,普遍存在的问题就是分类级别的局限性,一般一个分类指标只能适用于少数几个低级分类单位,不能对整个生物界做出总体上的系统分类分析。16S rRNA 二级结构则突破了这个局限,一方面,它在生物界广泛存在,并且早已经被证明是同源且保守的,非常适合作为系统分类学的研究工具;另一方面,核酸数据库日益庞大,测出的核酸碱基数呈指数增长,引入二级结构以后,同样的数据量能提供更多的系统发育信息,这将更加有利于系统分类学研究。相信必然会有更多更准确的预测方法被研究开发出来,16S rRNA 二级结构将越来越广泛的被用于系统分类研究。

#### 参考文献:

- [1] 刘海军, 史定华, 王翼飞. 日新月异的 RNA 二级结构预测[J]. 自然杂志, 2003, 25(6):314 - 322.
- [2] Noller H F, Woese C R. Secondary Structure of 16S Ribosomal RNA[J]. Science, 1981, 212:403 - 411.
- [3] Woese C R, Gutell R, Noller H F, et al. Detailed Analysis of the Higher-Order Structure of 16S-Like Ribosomal Ribonucleic Acids[J]. Microbiol Rev, 1983, 47(4):621 - 669.
- [4] Neefs J M, Van de Peer Y, De Rijk P, et al. Compilation of small ribosomal subunit RNA structures[J]. Nucleic Acids Res, 1993, 21(13):3 025 - 3 049.
- [5] Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information[J]. Nucleic Acids Res, 1981, 9(1):133 - 148.
- [6] Eddy S R, Durbin R. RNA sequence analysis using covariance models[J]. Nucleic Acids Res, 1994, 22(11):2 079 - 2 088.
- [7] Winker S, Overbeek R, Woese C R, et al. Structure detection through automated covariance search[J]. Comput Appl Biosci, 1990, 6(4):365 - 71.
- [8] James B D, Olsen G J, Pace N R. Phylogenetic comparative analysis of RNA secondary structure[J]. Meth Enzymol, 1989, 180:227 - 239.
- [9] Juan V, Wilson C. RNA Secondary Structure Prediction Based on Free Energy and Phylogenetic Analysis[J]. J Mol Biol, 1999, 289(4):935 - 947.
- [10] Cannone J J, Subramanian S, Schnare M N, et al. The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs[J]. BMC Bioinformatics, 2002, 3(1):2.
- [11] Pettersson B, de Silva SK, Uhlen M, et al. *Bacillus siralis* sp. nov., a novel species from silage with a higher order structural attribute in the 16S rRNA genes[J]. Int J Syst Evol Microbiol, 2000, 50:2 181 - 2 187.
- [12] Wisotzkey J D, Jurtshuk P Jr, Fox G E, et al. Comparative sequence analyses on the 16S rRNA (rDNA) of *Bacillus acidodardarius*, *Bacillus acidoterrestris*, and *Bacillus cycloheptanicus* and proposal for creation of a new genus, *Alicyclobacillus* gen. nov[J]. Int J Syst Bacteriol, 1992, 42:263 - 269.
- [13] Chen C, Zhao S, Ben K. Phylogenetic analysis of the family Thermaceae with an emphasis on signature position and secondary structure of 16S rRNA[J]. FEMS Microbiol Lett, 2003, 221(2):293 - 298.
- [14] Guo C L, Wang T, Peng Q, et al. The Phylotype of *Thermus* from the Rehai Geothermal Area, Tengchong, China[J]. The Journal of Microbiology, 2003, 41(2):152 - 156.
- [15] Rees G N, Harfoot C G, Janssen P H, et al. *Thiobaca trueperi* gen. nov., sp. nov., a phototrophic purple sulfur bacterium isolated from freshwater lake sediment[J]. Int J Syst Evol Microbiol, 2002, 52:671 - 678.
- [16] Hong S G, Bae K S, Herzberg M, et al. *Candida kunwiensis* sp. nov., a yeast associated with flowers and bumblebees[J]. Int J Syst Evol Microbiol, 2003, 53:367 - 372.
- [17] Sokolova T G, Gonzalez J M, Kostrikin N A, et al. *Thermosinus carboxydivorans* gen. nov., sp. nov., a new anaerobic, thermophilic, carbon monoxide-oxidizing, hydrogenogenic bacterium from a hot pool of Yellowstone National Park[J]. Int J Syst Evol Microbiol, 2004, 54:2 353 - 2 359.
- [18] Bouthinon D, Soldano H. A new method to predict the consensus secondary structure of a set of unaligned RNA sequences[J]. Bioinformatics, 1999, 15(10):785 - 798.
- [19] Akutsu T. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots[J]. Discrete Applied Mathematics, 2000, 104:45 - 62.