

# RNA 二级结构在微生物系统发育分析上的应用\*

刘 杨 崔晓龙 李文均 彭 谦\*\*

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

**摘要:** RNA 二级结构用于微生物分类和系统发育分析在近年来逐渐受到注意并发展起来。文章就此方面的发展做了简要综述, 其中对在系统发育分析上的应用部分作了较重点介绍。

**关键词:** RNA 二级结构, 分类, 系统发育, 根

**中图分类号:** Q93 **文献标识码:** A **文章编号:** 0253-2654 (2005) 02-0147-04

## Application of RNA Secondary Structure in Phylogenetic Analysis of Microbiology\*

LIU Yang CUI Xiao-Long LI Wen-Jun PENG Qian\*\*

(The Key Laboratory for Microbial Resources of Ministry of Education, Yunnan Institute of Microbiology, Yunnan University, Kunming 650091)

**Abstract:** Attention was gradually paid by biologists to the using of RNA secondary structure in the classification of microbiology and phylogenetic relationship analysis in recent years. The development around the research was summarized here briefly. And more emphasis was given to the part introducing the application of RNA secondary structure to the analysis of phylogenetic relationship.

**Key words:** RNA secondary structure, Classification, Phylogeny, Root

RNA 是生物生命活动中不可或缺的重要生物大分子。自核酶概念提出后, RNA 领域的新发现不断出现, 使人们对 RNA 在生命活动中所起作用的认识有了全新改变, 甚至令人们认为生命起始于 RNA 而非 DNA<sup>[1]</sup>。生物学家们试图通过对 RNA 结构的研究而找到它发挥作用的机制。在对 RNA 二级结构作了大量比较研究的基础上, 生物学家们发现 RNA (主要是指 ncRNA, 即非编码蛋白质 RNA, 下文同) 二级结构具有比一级结构更大的保守性, 且在 RNA 二级结构内发现了丰富的可用于分类和系统发育分析的信息。现在生物系统发育树的建立, 主要是构建在 SSU rRNA 及少数蛋白质序列比对的基础上, 它的建立对人们认识生物系统进化发育关系产生了深刻影响, 也极大促进了生物学, 特别是微生物学的发展。但是 SSU rRNA 序列分析方法客观上存在不足, 在一定程度上影响了它构建的树的准确性<sup>[2]</sup>。为了寻求更接近自然的生物系统进化发育关系, 生物学家们把眼光放到了各种具有进化保守性、普遍存在性及研究广泛性特点的生物分子上<sup>[3]</sup>。因为 RNA 二级结构比其一级结构在进化上具有更大的保守性, 基于一级结构的 RNA 二级结构预测也使得大量 RNA 二级结构易于获得, RNA 二级结构数据库也已经出现并收集保存了大量微生物的 RNA 二级结构, 因此 RNA 二级结构也成为分类和系统发育学家关注的对象。

\* 国家自然科学基金资助项目 (No. 39960003, 30360004)

Project Granted by Chinese National Natural Science Fund (No. 39960003, 30360004)

\*\* 通讯作者 Tel: 0871-5033543, E-mail: pengq2002@yahoo.com.cn

收稿日期: 2005-07-21, 修回日期: 2005-11-22

## 1 RNA 二级结构

和其它生物大分子一样, RNA 也存在着二级结构以及更高级结构。RNA 二级结构概念最早由 Doty<sup>[4]</sup> 和 Fresco<sup>[5]</sup> 提出, 它是由 RNA 单链自身回折而形成部分碱基配对和单链交替出现的茎环结构。相对于 RNA 序列, 在许多研究中, RNA 二级结构连同更高级结构被作为 RNA 的表型特征来描述, 以表明它与其基因型即 RNA 序列的不同。作为联系 RNA 序列与 RNA 更高级结构的中间结构, 它成了生物学家通过 RNA 序列了解 RNA 三级结构的一个门径, 进而了解结构与功能间的关系和 RNA 发挥其功能的机制<sup>[6]</sup>。现在通过物理化学方法对 RNA 结构的研究已经证明, RNA 二级结构中的许多环状结构与 RNA 功能密切相关。

## 2 RNA 二级结构的预测

RNA 二级结构应用于分类与进化研究的前提是能够得到准确的 RNA 二级结构图形。这点现在主要通过 RNA 二级结构的计算机预测解决。RNA 二级结构的预测最早应从 Tinoco 等<sup>[7]</sup> 对 RNA 二级结构相关能值估算开始, 到如今已经历 30 余年发展, 其间涌现了大量不同的计算机预测算法。根据它们原理的不同大致可分为 4 类: (1) 主要以最小自由能规则为依据的算法; (2) 利用点共变规则的比较序列算法; (3) 模仿分子自然进化过程来预测的遗传算法; (4) 以语言学相关理论来预测的随机上下文无关文法。四种算法中以第二种可靠性最强, 但是比较序列法手工操作较多, 大批量预测 RNA 结构时工作量较大; 最小自由能法在序列较短时能保证所预测结构的相对准确性 (预测 700bp 长的 RNA 片段时, 精确度达 70%), 而且操作相对简单, 可以直接用软件或提交到相关网站预测。

## 3 RNA 二级结构在微生物分类上的应用

分类学家曾在不同的分类水平上找到微生物 RNA 一级结构的特征性序列, 并成功地将其应用到微生物分类。RNA 二级结构在不同微生物中也存在特征性结构, 如独特的茎的长度或环的大小等。同时, RNA 二级结构还有一个特点, 即, 作为 RNA 序列在空间上折叠后的一种表现, RNA 二级结构能够突显出在序列比对中显得无足轻重, 实际上却比较重要的某一个或几个碱基突变。利用这样的特点, RNA 二级结构可以在不同的水平上对微生物进行很好的分类。如, Chen 等<sup>[8]</sup> 利用 16S rRNA 二级结构的 6, 10, 12 螺旋区结构特征信息对水生栖热菌进行了种的水平上的分类, 成功地将 38 株水生栖热菌划分到 18 个种内。Butler 等<sup>[9]</sup> 利用了 ribonuclease P RNA 二级结构特征将 *Gemmata* 群与其它细菌分开。Ishikawa 等<sup>[10]</sup> 利用 16S rRNA V6 区作为 *Marinilactibacillus* 区别于与之相近属的一个重要依据, 建立了一个新属。由于 RNA 二级结构比较直观, 又与细胞代谢及其它表型特征有着比一级结构更直接的联系, 近年来二级结构特征作为微生物的一个分类依据越来越受到微生物分类学家们的注意, 仅在国际微生物分类学权威杂志 IJSEM 上, 2002 年到现在不足 3 年时间内, RNA 二级结构相关文章就有近百篇之多, 其中又以 SSU rRNA 二级结构为主。RNA 二级结构特征, 现在在分类上主要还是作为一种参考指标而存在。但是从发展现状看, 它作为一种有效的分类依据, 已经得到学术界的普遍认可。

## 4 RNA 二级结构在系统发育分析上的应用

用 SSU rRNA 一级结构来研究与构建生物系统发育关系自 20 世纪 90 年代以来已经得到学术上的广泛认可。之所以选用 SSU rRNA 作为进化时钟是因为: (1) SSU rRNA 广泛存在于任何以细胞为基本结构单位的生物体内, 并且在进化上很保守; (2) SSU rRNA 相对较大 (约 1,500bp), 含有足够的进化信息; (3) SSU rRNA 序列已经大量人工测定, 很容易从数据库中获得。经多年研究发现, 除部分研究外, 大多数利用其它生物分子构建的系统树与用 SSU rRNA 序列构建的树大体一致, 从而在一定程度上肯定了 SSU rRNA 序列建树的可靠性<sup>[3]</sup>。但是也有其不足, 主要在于: (1) 难以确定根的位置 (现在所普遍认同的根, 是由少数蛋白质序列确定); (2) 客观存在的横向基因传递、碱基突变在时间与空间上的不平衡、高级结构对一级结构进化的束缚均使树的真实性受到很大影响<sup>[2]</sup>。为了弥补其中的不足, 使 SSU rRNA 序列建树更趋真实, 研究者把二级结构引入了建树过程, 认为参考二级结构信息可以评估与提高一级结构比对的准确性<sup>[11]</sup>。也有一些研究者试图直接利用 RNA 二级结构来研究与构建生物系统发育树。最早报道这方面的是 Billoud<sup>[12]</sup>, 他选择一类系统发育关系存在较大争议的海洋生物 Cirripede 作为研究对象, 把它们的 SSU rRNA 二级结构中无大量插入或缺失片断的部位分解成不同的亚结构单元, 每亚结构所含碱基数目作为其特征用数值描述, 利用所描述特征结合邻位相接法或最大简约法作出系统发育树。此方法在计算时, 并不考虑数值间差异的大小, 使所采用数据显得有些单一, 但是此方法及其中对一些问题 - 如不同结构碱基突变的权重问题等 - 的探讨仍值得系统发育学研究者借鉴。后来 Caetano-Anolles<sup>[13,14]</sup>又提出一种新的直接用各种 ncRNA 二级结构作树的方法。他采用数值字母法表示结构中不同特征性结构单元。用 0~9 和 A~P 表示不同结构单元中的碱基数目, 作为结构单元的编码。不同的结构单元按 5' - 3' 端方向排列, 就形成了一个由数字和字母组成的表示 RNA 二级结构特征的序列。用最大简约法对不同序列进行比对分析, 比对时参照结构从无序到有序、从热力学不稳定到稳定的进化过程, 最终预测出有根的系统发育树。研究中, 作者还选取了 RNA 功能位点的二级结构进行分析, 企图寻找出结构进化、功能进化与系统发育树间的关系。综合二者的研究我们可以发现, 直接用 RNA 二级结构来进行系统发育研究的关键在于如何采用合适的方式编码 RNA 二级结构, 把它转化为方便比对的线性形式。此外, 还必须要有适用于计算不同二级结构之间进化距离的算法, 这也是难点所在。

## 5 问题和展望

现在 RNA 二级结构在微生物分类与进化研究上的应用还处于探讨阶段。在微生物分类上, 它的应用范围正在逐渐扩大。但是, 是否任何分类单元间都存在特征性二级结构, 以及它可否作为一个独立的分类指标, 这些都有待更多的研究论证。

在系统发育分析上, 现在它主要作为一种参考来完善一级结构的比对, 直接拿它作为工具来用于研究的还不多。笔者认为原因有三: 首先, 对这方面的认识与注意是随着近年来对 RNA 结构研究的深入才逐渐有所增加; 其次, 二级结构远比一级结构复杂, 使得对其中一些问题 - 比如对确定结构特征比对时权重问题 - 的研究发展比较迟缓; 再者, 专门针对用二级结构来研究进化的软件尚未开发, 使得大量工作不得不手

工进行。但是,它的应用具有重要的潜在发展前景。正如国际知名杂志 *J Mol Evol* 编辑们<sup>[15]</sup>对 Caetano-Anolles 工作评价时所说,“这项工作的重要性不在于它的结果,而在于它的潜在发展前景。它代表着我们在寻求把结构和功能直接囊括到系统发育分析中的道路上迈出了重要一步。”

### 参 考 文 献

- [1] Kennedy D. *Science*, 2000, **290**: 2255 ~ 2257.
- [2] Doolittle W D. *Science*, 1999, **284**: 2124 ~ 2128.
- [3] Garrity M G. *Bergey's Manual of Systematic Bacteriology*. 2<sup>nd</sup> ed. Vol 1. Springer-Verlag New York Berlin Heidelberg, 2001. 49 ~ 65.
- [4] Doty P, Boedtker H, Fresco J R, *et al.* *J Proc of the Natl Acad of Sci of USA*, 1959, **45**: 482 ~ 499.
- [5] Fresco J R. *Nature*, 1960, **188**: 98 ~ 101.
- [6] Green R, Noller H F. *Annu Rev of Biochem*, 1997, **66**: 679 ~ 716.
- [7] Tinoco I Jr. *Nature*, 1971, **230**: 362 ~ 367.
- [8] Chen C Y, Zhao S L, Ben K L. *FEMS Microbiol Lett*, 2003, **221**: 293 ~ 298.
- [9] Butler M K, Fuerst J A. *Int J Syst Evol Microbiol*, 2004, **54**: 1333 ~ 1344.
- [10] Ishikawa M, Nakajima K, Yanagi M, *et al.* *Int J Syst Evol Microbiol*, 2003, **53**: 711 ~ 720.
- [11] Hickson R E, Simon C, Perrey S W. *Mol Biol Evol*, 2000, **17** (4): 530 ~ 539.
- [12] Billoud B, Guerrucci M A, Masselot M, *et al.* *Mol Biol Evol*, 2000, **17** (10): 1435 ~ 1445.
- [13] Caetano-Anolles G. *J Mol Evol*, 2002, **54**: 333 ~ 345.
- [14] Caetano-Anolles G. *Nucl Ac Res*, 2002, **30** (11): 2575 ~ 2587.
- [15] Pollock D D. *J Mol Evol*, 2003, **56**: 375 ~ 376.