

# 原生动物线粒体 DNA 的研究进展\*

张莉 钱雨 顾福康\*\*

(华东师范大学生命科学学院 上海 200062)

**摘要:** 目前已完成了多种原生动物的 mtDNA 全序列的测定,对线粒体基因组有了较全面和深入的认识。本文总结了原生动物的线粒体基因组结构、基因组成及基因表达等方面的研究进展,有助于进一步了解原始线粒体基因组的组成及其 mtDNA 进化,并为细胞的起源和真核生物进化的研究提供有价值的线索。

**关键词:** 原生动物的 mtDNA; 基因组结构; 基因组成; 基因表达

**中图分类号:** Q953 **文献标识码:** A **文章编号:** 0250-3263(2004)04-95-05

## Mitochondrial DNAs in Protozoa

ZHANG Li QIAN Yu GU Fu-Kang

(Department of Biology, East China Normal University, Shanghai 200062, China)

**Abstract:** In recent years many complete mtDNA sequences from protozoa have been revealed, which contributed a great deal to our knowledge of mtDNA. This paper reviews the research progress about protozoa mitochondrial genome structure, gene content, expression of certain genes etc. to make us better understand the original mitochondrial genome organization, mtDNA evolution, cell origin and the eukaryote evolution.

**Key words:** Protozoa mtDNA; Genome structure; Gene content; Gene expression

线粒体 DNA(mtDNA)是细胞内较小的复制单位,基因组结构较为简单,是研究 DNA 结构、基因组成、转录及表达的良好模型系统。并且利用对 mtDNA 多态性研究可为细胞起源分化,种内及种间系统发生关系提供证据,因此被广泛地作为系统发生和进化的标记。低等真核类的原生动物的具有比动物、植物、真菌等更丰富的生物多样性,它是这一系列研究中不可或缺的材料<sup>[1]</sup>。目前已对多种原生动物的 mtDNA 作了较详细的研究,其中包括 mtDNA 结构、基因组成、基因编码和部分基因表达等方面的研究。

### 1 线粒体基因组结构

原生动物的 mtDNA 呈线状或环状,其中以环状居多。已发表的原生动物线粒体基因组全序列测定工作表明,大多数原生动物的线粒体基因组大小在 15 ~ 70 kb 之间(表 1)。比较特殊的例子是顶复虫类(apicomplexan)线粒体,其基因组大小为 6 kb 左右(它们是已知的最小 mtDNA)<sup>[2,3]</sup>。从已经测序的各种原生动物的线粒体基因组来看,各基因间排列紧密,基本无大的非编码区,内

部间隔区也很少。整个基因组富含 A + T,尤其是在非编码区。与原生动物线粒体基因组相比,一般动物线粒体基因组通过剔除内部间隔区和缺失部分基因使基因组排列显得更为紧密;相反,植物线粒体基因组大小在 186 ~ 240 kb 之间,包含大量非编码区,仅很小区域具编码功能<sup>[4,5]</sup>。因此可以认为在真核生物进化产生高度趋异的线粒体基因组的过程中,原生动物的可能具有不同特征。

### 2 线粒体基因组组成

动物和真菌 mtDNA 基因组包含 13 种蛋白质编码基因(nad1 ~ 6, nad4L, cob, cox1 ~ 3, atp6, atp8),基因编码产物参与电子传递及偶联氧化磷酸化,此外动物和

\* 国家自然科学基金资助项目(No. 30270160);

\*\* 通讯作者, E-mail: fkgu@bio.ecnu.edu.cn;

第一作者介绍 张莉,女,24岁;研究方向:原生动物的细胞及分子生物学。

收稿日期:2003-10-20,修回日期:2004-04-10

表 1 原生动动物线粒体 DNA 结构特征<sup>[2,8]</sup>

名称	简称	形状	大小(bp)	编码区所占比例(%)	基因组 A + T 含量
<i>Acanthamoeba castellanii</i>	ACA	环形	41 591	93.2	70.6
<i>Chlamydomonas eugametos</i>	CEU	环形	22 897	84.6	65.4
<i>Chondrus crispus</i>	CCR	环形	25 836	94.8	72.1
<i>Dictyostelium discoideum</i>	DDI	环形	55 564	90.5	72.6
<i>Monosiga brevicollis</i>	MBR	环形	76 568	47	86
<i>Nephroselmis olivacea</i>	NOL	环形	45 223	78.4	67.2
<i>Paramecium aurelia</i>	PAU	线形	40 469	86.7	58.8
<i>Pedinomonas minor</i>	PMI	环形	25 137	60.9	77.8
<i>Plasmodium falciparum</i>	PFA	环形	5 966	76	68.4
<i>Porphyra purpurea</i>	PPU	环形	36 753	90.8	66.5
<i>Prototheca wickerhamii</i>	PWI	环形	55 328	70.6	74.2
<i>Reclinomonas americana</i>	RAM	线形	47 172	96	78.7
<i>Tetrahymena pyriformis</i>	TPY	环形	22 289	63.6	76.6

真菌 mtDNA 基因组还包含核糖体 rRNA 大小亚基基因 (rnl, rns)。植物线粒体基因组除上述基因之外,还有一些特殊的呼吸链蛋白质基因(如 nad7、nad9 等),且包含动物和真菌 mtDNA 中不存在的 5S rRNA 基因和核糖体蛋白质基因。序列比较分析表明,原生动动物线粒体基因组组成更接近于植物。Lang 等发现,异养鞭毛虫 (*Reclinomonas americana*) 的 mtDNA 包含所有线粒体基因组组成,此外,异养鞭毛虫 mtDNA 基因还有类似细菌类的  $\alpha$ -蛋白菌的特征(如 rnpB 基因编码 RNA 酶)。根据异养鞭毛虫 mtDNA 序列推断,它更接近于原始线粒体基因组,其它 mtDNA 基因只是异养鞭毛虫 mtDNA 的亚型,可能在进化过程中发生了不同程度的基因缺失(如真菌和动物共同的祖先发生了呼吸链蛋白质基因和核糖体蛋白质基因的缺失)<sup>[6]</sup>。尽管异养鞭毛虫 mtDNA 包含大量其它原生动动物 mtDNA 不具备的基因,但这些“多余”的基因都涉及到线粒体功能及起源。根据不断积累的原生动动物 mtDNA 全序列,现已能较准确地说明在线粒体基因组进化过程中基因遗失的次数和时间,其中很多基因遗失现象涉及到线粒体基因向核基因的转移<sup>[7]</sup>。

**2.1 蛋白质编码基因** 如前所述,原生动动物 mtDNA 基因组含有多种蛋白质编码基因,包括 NADH 氧化还原酶亚基基因 (nad1 ~ 6) (部分原生动动物 mtDNA 还包括 nad7 ~ 11 基因)、细胞色素氧化酶三个亚基基因 (cox1 ~

3)、细胞色素 b 脱辅基蛋白基因 (cob)、ATP 酶亚基基因 (atp1, 3, 6, 8, 9)。另外红藻 (*Chondrus crispus*、*Porphyra purpurea*) 和异养鞭毛虫还包含琥珀酸脱氢酶亚基基因 (sdh2 ~ 4)。

大多数原生动动物含有大量保守但功能未知的开放阅读框(ORF)。已初步确定原生动动物线粒体 orfB 基因编码 ATP 合成酶的 F<sub>0</sub> 部分亚基,与植物和真菌的 atp8 基因同源<sup>[8]</sup>。近来对鞭毛虫狭叶滴虫 (*Seculamonas ecuadoriensis*) mtDNA ymf39 基因序列及 ATP 酶部分蛋白质序列分析证明其 ymf39 与真菌 atpF 基因同源,编码 F<sub>0</sub>F<sub>1</sub> 型 ATP 合成酶亚基 b。还发现 Ymf39 与核基因编码蛋白 ATP4 功能相似,当原生动动物 mtDNA 中无 ymf39 基因时在核基因组中存在 atp4 基因。推测,atp4 基因可能来源于线粒体基因组<sup>[9]</sup>。大部分原生动动物 mtDNA ORF 基因序列特殊,目前尚不能确定其对应的蛋白质序列。

**2.2 rRNA** 如表 2 所显示,几乎所有原生动动物线粒体基因组都包括两个 rRNA 基因 (rnl, rns),编码两种大小 rRNA (LSU rRNA, SSU rRNA),它们的二级结构一般很保守。大部分原生动动物线粒体 LSU rRNA 5' 端和 3' 端区域结构类似于细菌如大肠杆菌 (*Escherichia coli*) 23S rRNA 的相应部分结构<sup>[10]</sup>。而多细胞动物线粒体 LSU rRNA 则不存在类似的结构特征。这些研究进一步说明了原始线粒体基因组存在于原生动动物中的可能性。

表 2 原生动动物 mtDNA rRNA 基因<sup>[8,12,14]</sup>

rRNA	ACA	CEU	CCR	DDI	MBR	NOL	PAU	PMI	PFA	PPU	PWI	RAM	TBR	TPY
rnl	√	√	√	√	√	√	√ <sup>a</sup>	√ <sup>a</sup>	√	√	√	√	√	√ <sup>a</sup>
rns	√	√	√	√	√	√	√ <sup>b</sup>	√	√	√	√	√	√	√ <sup>b</sup>
rrn5	√	-	-	-	-	-	-	-	-	-	√	-	-	-

√表示基因存在; - 表示基因不存在。a 表示 rnl 为断裂重排基因; b 表示 rns 为断裂重排基因

有一部分原生动动物 mtDNA 的 rRNA 基因异常,如锥虫 (*Trypanosoma brucei*) 和利什曼虫 (*Leishmania tarentolae*) 的 9S rRNA (SSU rRNA) 和 12S rRNA (LSU rRNA) 基因,其 rRNA 二级结构高度变异,仅包含一部分已知的保守序列<sup>[11]</sup>;梨形四膜虫 (*Tetrahymena pyriformis*), 双小核草履虫 (*Paramecium aurelia*) 和小平滴虫 (*Pedinomonas minor*) 两个 rRNA 基因都为断裂重排基因,分别包括两个编码区 ml\_a, ml\_b 和 ms\_a, ms\_b<sup>[12]</sup>。

另外与动物及真菌 mtDNA 一样的是大部分原生动动物 mtDNA 缺少 5S rRNA 基因 (rm5)。但衣绿藻 (*Prototheca wickerhamii*) 和异养鞭毛虫线粒体基因组中有 5S rRNA 基因,其中包括一高度保守区和大多数的可变区,基因组序列和二级结构与细菌及植物 5S rRNA 基因相似<sup>[6,13]</sup>。曾经认为红藻线粒体基因组 *cox2* ~ *cox3* 基因间隔区中也包含 5S rRNA 基因<sup>[14]</sup>,然而研究红藻基因组序列发现:这个假定的 5S rRNA 基因序列前 42 个核苷酸序列与 *cox2* 基因 3' 端编码区重叠,且它的 5S rRNA 基因序列缺少完全高度保守区,其二级结构与植物及 *P. wickerhamii* 和 *R. americana* 的 5S rRNA 二级结构也不完全符合。在红藻 *P. purpurea*、*Gracilaria lemaneiformis* mtDNA 中也存在类似情况。由此推断,这些假定的 5S rRNA 基因仅是非功能性的假基因。比较已知的真核生物 mtDNA 序列,5S rRNA 基因在真核生物中呈无规则分布,因此有人推测如果 5S rRNA 基因是单系起源,那么可能在真核生物进化过程中发生了多次 rm5 基因的缺失<sup>[15]</sup>。

**2.3 tRNA** 一般原生动动物线粒体基因组编码 tRNA 仅能识别 mRNA 链上一部分密码子,不能满足线粒体蛋白质合成需要,典型的例子是卡氏棘变形虫 (*Acanthamoeba castellanii*), 盘基网柄虫 (*Dictyostelium discoideum*), 以及双小核草履虫和梨形四膜虫,它们的 mtDNA 编码 tRNA 数量远少于翻译过程中所需 tRNA。

另外在原生动动物顶复虫类和锥虫类线粒体基因组中至今未发现 tRNA 基因。已研究证明利什曼虫参与线粒体蛋白翻译的 tRNA 都是由核基因组编码,在线粒体外合成加工形成成熟的 tRNA 后转移到线粒体内行使功能的位置上<sup>[16]</sup>。目前所知大部分原生动动物线粒体基因组都能编码一些特定的 tRNA 基因,如 tRNA<sup>Met</sup>、tRNA<sup>Trp</sup>, 而其它 tRNA 基因 (特别是 tRNA<sup>Thr</sup> 基因) 在原生动动物中并不常见。

原生动动物线粒体基因组编码 tRNA 二级结构一般为传统的三叶草型结构。比较特殊的例子是四膜虫线粒体中 tRNA<sup>Met</sup>, 其二级结构中 D 茎环部分较短,反密码子环中有两个相邻的假尿嘧啶 ( $\Psi$ )<sup>[17]</sup>。这些特征可能与其能识别不同的起始密码子相关联<sup>[12]</sup>。除此之外,四膜虫 tRNA<sup>Met</sup> 还能形成 L 型三级结构。

**2.4 内含子** 根据内含子的保守序列及二级结构,内含子可分为 I 类内含子和 II 类内含子。原生动动物 mtDNA 与植物 mtDNA 相比,内含子数量非常少。在已完成测序的原生动动物 mtDNA 中 (表 3), 有近一半的原生动动物完全缺少 I 类和 II 类内含子。研究发现卡氏棘变形虫、盘基网柄虫、衣绿藻、橄榄肾形藻虫 (*Nephroselmis olivacea*) 和真配子衣滴虫 (*Chlamydomonas eugametos*) 及领鞭毛虫 (*Monosiga brevicollis*) 中存在 I 类内含子。*P. wickerhamii* mtDNA *cox1* I 类内含子与植物和真菌 mtDNA 对应内含子位置相同,结构同源。这些内含子可能都起源于绿藻、植物及真菌共同的线粒体祖先<sup>[18]</sup>。另外发现卡氏棘变形虫 mtDNA *ml* 基因中 I 类内含子和一些衣滴虫叶绿体 DNA 的 I 类内含子不仅位置相同且具有同源的内含子核心结构及内含子 ORF,说明可能发生了 I 类内含子水平转移<sup>[19]</sup>。原生动动物 mtDNA 中 II 类内含子所占比例则更少。有证据表明灰藻 *Pylaiella littoralis* mtDNA *ml* 基因是通过内含子水平转移得到 II 类内含子的<sup>[20]</sup>。

表 3 原生动动物线粒体内含子数<sup>[8,18]</sup>

	ACA	CEU	CCR	DDI	MBR	NOL	PAU	PMI	PFA	PPU	PWI	RAM	TBR	TPY
I 类内含子	3	9	0	5	4	4	0	0	0	0	2	0	0	0
II 类内含子	0	0	1	0	0	0	0	1	0	2	0	1	0	0

### 3 线粒体蛋白质基因表达

虽然关于线粒体基因组结构和序列分析的工作逐渐增多,但对线粒体基因表达的研究仍较少。目前,原生动动物中仅对四膜虫线粒体蛋白质基因表达体系有了较详细的了解。这些结果使我们对原生动动物线粒体的转录、RNA 加工翻译体系有了一定的认识,也为原生动

物线粒体基因表达机制的多样性提供了例证。

**3.1 四膜虫线粒体 DNA 转录** 四膜虫线粒体 mRNA 多为单顺反子,极少数为双顺反子或多顺反子。其中 *nad7* 基因 mRNA 为多顺反子。另外 *Nad7*、*Rps14* 和 *Ymf60* 蛋白由同一顺反子翻译而来。一般认为在四膜虫线粒体基因组中只有很少的转录起始位点,从这些位点起始转录产生较长的前体 mRNA,然后加工成短的

单顺反子 mRNA<sup>[21]</sup>。

**3.2 nad1 基因转录** 四膜虫 mtDNA 中存在一特殊 nad1 基因,它由两个不同的编码区:nad1\_a, nad1\_b。位于 mtDNA 相反链上<sup>[17]</sup>。Northern 杂交实验检测到与 nad1\_a 和 nad1\_b 基因大小相同和相对较长但未充分转录的基因,但未发现有与 nad1 基因大小一致的 mRNA<sup>[21]</sup>。这表明在四膜虫线粒体转录体系中没有通过反式拼接来产生共价连续的可翻译的 mRNA。

**3.3 起始密码子及转录产物 5' 端** 四膜虫 mtDNA 中起始密码子 AUG 仅存在于一部分 ORF 起始处,其余的 ORF 中或缺乏起始密码子 AUG,或 AUG 位于终止密码子 UAA 下游,甚至整个编码区无起始密码子 AUG。这表明在蛋白质编码过程中可能使用了非标准起始密码子或通过 RNA 编辑形成标准的起始密码子。对四膜虫线粒体基因组研究确定,四膜虫线粒体 mRNA 翻译系统既可使用标准起始密码子 AUG,还可以利用非标准起始密码子 AUU、AUA、GUG 和 GGG<sup>[21]</sup>。nad2 和 nad7 基因是四膜虫线粒体基因组中两个特殊基因,它们在基因组中位置相邻但转录方向相反。在各自的 ORF 起始处都有密码子 AUG,且两个 ORF 有约 90 个密码子大小的重叠区,若从 AUG 起始转录,将产生 N 端延伸的特异蛋白,而使用其它起始密码子则解决了这一问题。

这些起始密码子位于转录产物 5' 端 2~5 个碱基处,所以四膜虫线粒体 mRNA 完全缺少 5' 前导序列。一般 5' 前导序列能形成茎环结构降低翻译水平或抑制蛋白结合 5' 端阻止 mRNA 翻译。因此无这种 5' 前导序列有利于四膜虫起始蛋白质翻译。

## 4 展 望

从 1992 年 OGMP (*organelle genome megasequencing program*) 实施以来,已有很多原生动物的 mtDNA 完成了全序列的测定,对原生动物的线粒体基因组的结构、组成正逐步形成系统的认识。过去,由于植物和衣滴虫 mtDNA 结构组成及表达方式很不同,似乎很少有证据表明它们有共同的进化起源,曾认为植物线粒体基因组祖先较其它真核生物线粒体基因组进化祖先更近<sup>[22]</sup>。然而衣绿藻和其它原生动物的线粒体基因组的测定表明植物 mtDNA 与原生动物的 mtDNA 有共同的进化祖先<sup>[23,24]</sup>。到目前为止仍有相当一部分原生动物的 mtDNA 序列有待测定,对它们的了解将进一步完善我们对原始线粒体基因组组成的认识,进而为细胞起源、生物进化提供又价值的线索。由于线粒体是一个半自主的细胞器,线粒体基因组与核基因组之间存在千丝万缕的联系。本实验室已有的研究表明冠突伪尾柱虫

(*Pseudourostyla cristata*) 休眠细胞由于外界不良环境的影响,细胞内发生一系列连锁反应,最终可能会关闭某些核基因组内表达线粒体功能组分的基因或激活核基因组中某些基因,从而合成某种特定因子,去阻断线粒体 DNA 的表达甚至主动破坏线粒体 DNA,使线粒体 DNA 不能正常表达。在原生动物中关于核基因组对线粒体基因如何表达调控、及线粒体基因组对核基因组如何反馈调节的详细过程还不太清楚。近来研究发现植物中不同的核背景会导致线粒体基因组结构的改变,植物的一些线粒体基因以 DNA 而非 RNA、cDNA 或整个细胞器染色体形式进入核基因组<sup>[25]</sup>,在原生动物中是否也存在这样的情况? 这些是今后对原生动物的 mtDNA 的研究方向之一,将有助于对真核生物有更为全面的了解。

## 参 考 文 献

- [1] Burger G, Lang B F. Parallels in genome evolution in mitochondrial and bacterial symbions. *IUBMB Life*, 2003, **55**: 205 ~ 212.
- [2] Feagin J E. The extrachromosomal DNAs of apicomplexan parasites. *Annu Rev Microbiol*, 1994, **48**: 81 ~ 104.
- [3] Sharma I, Pasha S T, et al. Complete nucleotide sequence of the *Plasmodium vivax* 6 kb element. *Mol Biochem Parasitol*, 1998, **97**: 259 ~ 263.
- [4] Boore J C. Animal mitochondrial genomes. *Nucleic Acids Res*, 1999, **27**: 1 767 ~ 1 780.
- [5] Hanson M R, Folkerts O. Structure and function of the higher plant mitochondrial genome. *Int Rev Cytol*, 1992, **141**: 129 ~ 172.
- [6] Emelyanov V V. Mitochondrial connection to the origin of the eukaryotic cell. *Eur J Biochem*, 2003, **270**: 1 599 ~ 1 618.
- [7] Covello P S, Gray M W. Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (cox2) in soybean: evidence for RNA-mediated gene transfer. *EMBO J*, 1992, **11**: 3 815 ~ 3 820.
- [8] Gray M W, Lang B F, Cedergren R, et al. Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res*, 1998, **26**: 865 ~ 878.
- [9] Burger G, Lang B F, Braun H P, et al. The enigmatic mitochondrial ORF ymf39 codes for ATP synthase chain b. *Nucleic Acids Res*, 2003, **31**: 2 353 ~ 2 360.
- [10] Gutell R R, Gray M W, Schnare M W. A compilation of large subunit (23S and 23S-like) ribosomal RNA structures. *Nucleic Acids Res*, 1993, **21**: 3 055 ~ 3 074.
- [11] Gutell R R. Collection of small subunit (16S-and 16S-like) ribosomal RNA structures. *Nucleic Acids Res*, 1994, **22**: 3 502

- ~ 3 507.
- [12] Burger G, Zhu Y, Littlejohn T G, *et al.* Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. *J Mol Biol*, 2000, **297**:365 ~ 380.
- [13] Wolff G F, Plante I, Lang B F, *et al.* Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*. *J Mol Biol*, 1994, **237**:75 ~ 76.
- [14] Leblanc C, Boyen C, Richard O, *et al.* Complete sequence of the rhodophyte *Chondrus crispus*. *J Mol Biol*, 1995, **250**:484 ~ 495.
- [15] Lang B F, Goff L J, Gray M W. A 5S rRNA is present in the mitochondrial of the protist *Reclinomonas americana* but is absent from red algal mitochondrial DNA. *J Mol Biol*, 1996, **261**:607 ~ 603.
- [16] Kapushoc S T, Alfonso J D, Rubio M A, *et al.* End processing precedes mitochondrial importation and editing of tRNAs in *Leishmania tarentole*. *J Biochem*, 2000, **275**:37 907 ~ 37 914.
- [17] Brunk C F, Lee L C, Tran A B, *et al.* Complete Sequence of the mitochondrial genome of *Tetrahymena thermophila* and comparative methods for identifying highly divergent genes. *Nucleic Acids Res*, 2003, **31**:1 673 ~ 1 682.
- [18] Wolff G F, Burger G, Lang B F, *et al.* Mitochondrial genes in the colourless alga *Prototheca wickerhamii* resemble plant genes in their exons but fungal genes in their introns. *Nucleic Acids Res*, 1993, **21**:719 ~ 726.
- [19] Turmel M, Cote V, Otis C, *et al.* Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion). *Mol Biol Evol*, 1995, **12**:533 ~ 545.
- [20] Fontaine J M, Rousvoal S, Leblanc C, *et al.* The mitochondrial LSU rDNA of the brown alga *Pylaiella littoralis* reveals alpha-proteobacterial features and is split by four group IIB introns with an atypical phylogeny. *J Mol Biol*, 1995, **251**:378 ~ 389.
- [21] Edqvist J, Burger G, Gray M W. Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. *J Mol Biol*, 2000, **297**:381 ~ 393.
- [22] Gray M W, Cedergren R, Abel Y, *et al.* On the evolutionary origin of the plant mitochondrion and its genome. *Proc Natl Acad Sci USA*, 1989, **86**:2 267 ~ 2 271.
- [23] Burger G, Plante I, Lonergan K M, *et al.* The mitochondrial DNA of the amoeboid protozoon, *Acanthamoeba castellanii*. Complete sequence, gene content and genome organization. *J Mol Biol*, 1995, **245**:522 ~ 537.
- [24] Gray M W, Burger G, Lang B F. The origin and early evolution of mitochondria. *Genome Biology*, 2001, **2**:10 181 ~ 10 185.
- [25] Henze K, Martin W. How mitochondrial genes get into the nucleus? *TRENDS in Genetics*, 2001, **17**: 383 ~ 387.