纤毛虫原生动物的分子生物学研究: 若干热点领域及新进展

伊珍珍^{①†}, 苗苗^{②†}, 高珊^{③†}, 高凤^{③†}, 宋微波^{③*}

① 华南师范大学广州亚热带生物多样性与环境生物监测重点实验室, 广州 510631;

② 中国科学院大学存济医学院, 北京 100049;

③ 中国海洋大学海洋生物多样性与进化研究所, 青岛 266003

† 同等贡献

* 联系人, E-mail: wsong@ouc.edu.cn

2016-01-16 收稿, 2016-03-30 修回, 2016-03-31 接受, 2016-06-07 网络版发表

青岛市应用基础研究计划(15-12-1-1-jch)和国家自然科学基金(31471973)资助

摘要 作为分子生物学研究的重要模式材料, 纤毛门原动物在细胞生物学、遗传学、基因组学、生态学以及真核生物的起源和进化等领域均具有重要的研究价值. 近年来, 本团队在纤毛虫多样性和细胞学研究的同时, 聚焦在纤毛虫的分子生物学领域并形成了一批新成果: 完成了多个代表性类群2000余条标记性基因的测序和提交以及基因标记筛选、拓扑结构优化等相关的方法学探讨; 以此为基础, 开展了门内50余个纲目级阶元系统关系的梳理以及基于多基因分析对整个纤毛门的系统树重建; 揭示了四膜虫(*Tetrahymena*)的一个重要表观遗传学新现象, 即特异性修饰组蛋白H3第27位赖氨酸的单甲基化酶TXR1的敲除将严重影响DNA复制延伸的速度和保守度, 证实了表观遗传信息的微扰可直接影响遗传信息的复制; 对沟缘游虫(*Strombidium sulcatum*)等6个代表种的转录组/全基因组/线粒体基因组的测序和分析, 为纤毛虫进化过程中关键节点种类定位、寻找功能基因家族与形态特征的相关性提供了大量信息; 首次发现了选择性拼接在钩刺斜管虫(*Chilodonella uncinata*)的基因家族中广泛存在并提出一新的理论模型: 选择性拼接可能是基因乱序进化过程的中间形式, 为探讨纤毛虫物种分化、形成机制提供了一新的视角; 在对DNA条形码候选基因分析比较的基础上, 初步揭示出不同纤毛虫类群生境间的迁移机制, 发现和定位了4个优良候选区段并构建了80余种代表性类群的基因条形码数据库.

关键词 纤毛门原动物, 分子系统学, 基因条形码, 表观遗传学, 基因组学, 进化

在漫长的动物进化过程中, 处于真核生物起源阶段的原生生物是一大类种类繁多、多样性极高、结构与功能均十分复杂的单细胞生物. 其中, 纤毛虫处于细胞器结构演化、分化的最高阶段: 无论在有性生殖方式、双态核型(功能、形态、倍性、数量迥异的大、小核)、还是极端多样化的分裂期亚结构模式形成与调控过程等, 均代表了细胞分化-进化的极高级

阶段. 作为重要的模式生物, 纤毛虫在细胞学、遗传学、生态学、进化生物学等领域都具有重要的研究价值. 例如, 以四膜虫(*Tetrahymena*)为模式的研究, 成就了近20年来包括揭示核酶、端粒酶及反式剪切等一系列生物学领域的重大发现^[1]. 尤其是近二、三十年来分子技术和手段的革命, 围绕纤毛虫为材料所开展的组学、表观遗传学、细胞结构的分化与去分化、

引用格式: 伊珍珍, 苗苗, 高珊, 等. 纤毛虫原动植物的分子生物学研究: 若干热点领域及新进展. 科学通报, 2016, 61: 2227-2238

Yi Z Z, Miao M, Gao S, et al. On molecular biology of ciliated protozoa: Frontier areas and progresses (in Chinese). Chin Sci Bull, 2016, 61: 2227-2238, doi: 10.1360/N972016-00064

模式调控、共生与协调进化以及分子系统发育等分支的研究,已成为当今国际原生动物学领域最为活跃热点方向之一。

在过去的30年中,本团队以纤毛虫原生动物为研究对象,系统地开展了对我国沿海纤毛虫的多样性与系统学、细胞学、生态学等领域的研究^[2]。作为近年来的新焦点,本团队在分子系统发育、基因条码技术探索、以四膜虫为材料的表观遗传学、大小核基因组重组以及纤毛虫组学等领域开展了多角度的探究,形成了一批新成果。本文围绕上述分支领域的进展及未来发展预期概述如下。

1 纤毛虫分子系统学

长期以来,对纤毛门原生动物系统关系的探讨除在纲目级等高阶阶元上参考了细胞分裂等个体发育信息外,大部分类群的系统发育关系主要是建立在纤毛器特化水平、纤毛图式、细胞核及超微结构等形态学特征的对比基础上^[3]。20世纪后期开始,利用标记性基因来探讨系统发育关系,为生物的自然系统构建开辟了一条新途径并迅速发展为主流手段^[4,5]。相对于后生动物,纤毛虫的分子系统学研究所遇到的困难在于,太多类群的分类学研究严重不足,很多进化节点类群分子信息缺失,由此制约了系统学分析的开展。

从20世纪末开始,作者所在原生动物学团队借助于对我国温带-热带海区所开展的纤毛虫区系分类学研究,完成了大量素材的DNA样本库的构建和以核糖体基因家族为代表的标记性基因的测序工作。在此基础上,对众多系统关系缺乏了解的类群开展了多层次的探讨,并逐步成为国际间该领域研究的核心团队之一:根据对Web of Science数据库的最新统计显示,在近10余年的国际主流刊物上,本团队与合作者所发表的本领域研究文章(129篇)占同期全球同类工作(916篇)的14%。迄今为止,先后开展了对腹毛类、游仆类、寡毛类、盾纤类、管口类等9个大类群的系统学研究(图1),覆盖了纤毛门内约2/3的目级阶元,系列成果连续在*Molecular Phylogenetics and Evolution*、*Proceedings of the Royal Society B*等领域主流期刊上发表了13篇研究文章。主要包括:(i)完成了包含1700余种/种群的海洋纤毛虫DNA库的构建,该库收集了几乎全部海洋中自由生活的常见类群;同步完成了对大量种-种群核糖体家族基因为主的测序工

作,涵盖了50%以上的目级阶元(32/59),在腹毛类等主要类群中,本团队所测序列占国际GenBank中已登录数据的50%以上(表1);(ii)在多基因分析的基础上,完成了对纤毛门大系统的构建和对门下50余目级阶元系统发育关系的厘定^[6];(iii)开展了对旋唇纲、盾纤亚纲、原腹毛亚纲、寡毛亚纲等6个纲级阶元内部关系的梳理,在此基础上建立了3个新纲、6个新目/亚目、3个新科^[7-9];(iv)工作还包括对纤毛虫系统树构建等方法学的探索:提出和深入探讨了在纤毛虫分子树构建中如何利用GenBank中大量环境未定种序列来提高系统树拓扑结构的可信度问题;评估了二级结构在分子系统发育研究中的应用价值;探讨了蛋白基因中旁系同源现象对纤毛虫分子系统树构建以及多基因数据集中的数据缺失和异质性的对系统树拓扑结构和置信值的影响^[10-13]。上述工作丰富了人们对纤毛虫系统发育与进化领域的认知,对后续的基因标记筛选、更全面的系统构建等研究的开展奠定了重要基础。

由于单细胞生物样品采集过程中存在的不可视性、不同类群间研究工作进展的不平衡性以及DNA提取受限于能否建立成功的培养等因素,纤毛虫分子系统发育仍有许多问题有待解决。例如,现在已完成至少一条基因测序的种类尚不足已知种的1/10;迄今对大部分类群所开展的分析仍主要(或仅)是基于小亚基核糖体rRNA单一基因信息;由于稳定度的差异,很多已有的工作包含了众多的误导和可信度不足问题等。一个在学术界普遍认可的观点是,未来的研究中,需要特别加强基于大样本、高覆盖度、组学及多基因、个体发育-形态学等综合信息对系统地位不明类群的探讨。此外,筛选新的具广适性的标记性基因以及完善目前GenBank中缺失类群的数据,依然是摆在系统学者面前最迫切的任务之一。

2 纤毛虫DNA条形码技术

自20世纪末以来,随着DNA指纹图谱、克隆文库、高通量测序等分子技术的应用,极大地加深了人们对原生动物生物多样性的认知^[14]。长期以来,生态学研究中对物种的鉴定高度依赖于分类学者的知识和经验,而DNA条形码技术的建立为基于基因序列的物种刻画提供了一条重要的新途径,即物种鉴定无需经由耗时、繁琐的过程和对操作者分类学背景的依赖。然而,与分子生态调查中产生的爆炸式增长

图1 (网络版彩色)纤毛虫分子系统学研究。(a) 腹毛类的18SrRNA基因树;(b) 叶咽纲的18SrRNA基因树;(c) 盾纤类的系统发育;(d) 18SrRNA基因序列中不同类群的特性位点

Figure 1 (Color online)Molecular phylogeny of ciliates. (a) Phylogenetic relationships within the core hypotrichs inferred from 18SrRNA gene sequences; (b) phylogenetic relationships within class Phyllopharyngea inferred from 18SrRNA gene sequences; (c) evolutionary scheme within scuticociliates; (d) group-specific nucleotide sites in 18SrRNA gene sequences

的环境未定种序列(operational taxonomic unit, OTU)相比,纤毛虫原生动物DNA条形码研究却进展缓慢^[15]。尽管DNA条形码概念提出已有10余年,纤毛虫以及其他原生动物类群的DNA条形码研究仍处于基因标记筛选阶段,目前仅有为数不多的关于纤毛虫DNA条形码研究的报导且集中于四膜虫等少数模式种类^[16-21]。

在纤毛虫的初期工作中,通常采用线粒体细胞色素c氧化酶I(cytochrome c oxidase subunit I, *COI*)

基因5'端的序列作为DNA条形码标记基因^[18,19]。但是随后的研究发现,该基因在不同类群中变异很大并经常存在较大的插入区,以至于无法设计通用引物并进而提出了大亚基核糖体(28SrRNA)基因5'端可能更为适合^[21,22]。本团队及合作者在前人工作的基础上,以前口虫属为研究对象,综合比较了5种基因8个不同区段的区分效力,包括线粒体*COI*基因和4种大核基因:18SrRNA基因V2, V4区,28SrRNA基因D1, D2区,转录间隔区1, 2(ITS1, 2), 5.8S核糖体RNA

表1 本团队所完成的主要纤毛虫类群核糖体rRNA基因测序情况

Table 1 Ciliate sequences: contribution to the GenBank rRNA gene database by authors' groups

纤毛虫类群	18S rRNA 基因序列条目			ITS-5.8S rRNA基因序列条目			28S rRNA基因序列条目		
	GenBank	本团队	贡献率(%)	GenBank	本团队	贡献率(%)	GenBank	本团队	贡献率(%)
直口纲(Litostomatea)	358	54	15	127	6	5	17	14	82
腹毛类(Hypotrichia & Euplotia)	705	207	29	286	97	34	113	87	77
寡毛类(Oligotrichia s.l.)	933	43	5	528	39	7	24	24	100
叶咽类(Phyllopharyngea)	140	76	54	9	4	44	7	6	86
篮管类(Nassophorea)	18	10	56	4	4	100	5	5	100
前口类(Prostomatea)	54	10	19	4	4	100	5	5	100
核残迹类(Karyorelictea)	59	27	46	3	2	67	1	1	100
异毛类(Heterotrichea)	179	24	13	83	7	8	8	8	100
缘毛类(Peritrichia)	261	20	8	230	14	6	40	6	15
盾纤类(Scuticociliatia)	206	78	38	89	49	55	35	35	100
咽膜类(Peniculia)	154	20	13	507	3	1	3	2	67

(5.8SrRNA) 基因。结果发现, *COI*, ITS1, ITS2, 28SrRNA基因D2区均能够很好地分开不同的种, 显示为良好的候选标记基因; 而18SrRNA基因V2, V4区、28SrRNA基因D1区和5.8SrRNA基因中均未检测到条形码间隙^[23]。此外, 探讨了不同层面的差异对种类鉴定和种群结构分析的影响, 明确了线粒体基因的异质性, 揭示了*COI*基因在同一细胞内部以及个体间、种群间的遗传差异(图2)^[24]。

在DNA条形码研究领域, 与基因标记筛选相比, 更为迫切的工作是对应于物种的数据库构建。而制约建库的首要因素是物种的采集与正确鉴定, 这需要专业的形态分类学知识, 为此, 通常物种鉴定人员需要经过长达数年或更长时间的系统培训^[25]。得益于本团队过去20多年中分类、区系研究工作的积累和由此构建的内容丰富的海洋纤毛虫DNA库, 为条形码研究的开展提供了丰富的素材。在此基础上, 近年来成功地获得了咽膜亚纲、寡毛亚纲、游仆亚纲等类群中累计83种的DNA条形码序列^[23,24], 涉及更广类群的工作也在顺利进展中。

越来越多的工作显示, 在分子生态学研究中被广泛应用的、几乎是目前唯一选择的18S基因并不适于纤毛虫的物种刻画或种间区分^[23,26]。在原生动物学界, 人们目前提出一个权宜的方案, 即通过“两步走”的方式去探讨条码技术应用的可行性: 先借助18SrRNA基因来形成初步的判读, 再根据需要采用更高变的基因完成进一步的刻画^[15]。毫无疑问, 在未来的工作中, 摆在原生动物学家面前的任务仍将

是深化对更多类群的了解、资料库的构建以及寻找不同类群中稳定共享的条形码间隙阈值, 从而为完善纤毛虫DNA条形码及宏条形码技术在多样性研究中的应用奠定基础。

3 纤毛虫组学研究

进入21世纪以来, 组学作为重大的革命性技术, 利用大量数据的整合来探究生命体内基因、蛋白质和生物大分子间的相互作用, 为研究生命的起源、发育、遗传、代谢等一系列生物学基本问题提供了新的视野, 并成为当前生物科学领域中最具影响力的热点研究领域^[27~29]。在最早启动的基因组计划中, 以纤毛虫原生动物——嗜热四膜虫为真核模式生物的大核基因组于2006年完成了测序、拼接、组装、注释及相关分析^[30]。随后, 累计已有12种纤毛虫完成了基因组测定, 超过20种完成了转录组的测定, 大量后续工作, 包括组装、注释、分析、新类群测序等也在计划或进行中^[27~29]。

在国内, 纤毛虫组学研究开展得较晚但呈现了明显的后发优势^[31~33]。出于研究意义的考虑, 本团队与合作者^[31~33]近年来主要开展了对6个具类群代表性的物种: 具沟急游虫(*Strombidium sulcatum*)、海洋尾丝虫(*Uronema marinum*)、相似后阿脑虫(*Metanophrys similis*)、扇形游仆虫(*Euplotes vannus*)、柠檬类瘦尾虫(*Uroleptopsis citrina*)、尾草履虫(*Paramecium caudatum*)的转录组测定, 结果包括: 初步完成了对急游虫转录组特征分析, 并基于现有的纤毛虫组学数据, 构建了

图2 (网络版彩色) 纤毛虫DNA条形码研究. (a) 基于DNA条形码序列和形态特征对前口虫的鉴别; (b, c) 线粒体 COI 基因异质性对草履虫种群结构的影响; (d) 线粒体 COI 基因在个体内、个体间、种群间的变异位点

Figure 2 (Color online) DNA barcoding of ciliates. (a) Species discrimination of the genus *Frontonia* based on DNA barcoding sequences and morphological characters; (b, c) population structures of *Paramecium* species affected by heteroplasmy of mitochondrial COI gene; (d) variable nucleotide sites of mitochondrial COI gene detected in intra-cell, inter-cell and inter-population levels

纲级水平的(系统发育)拓扑结构(图3). 研究显示, 表层胞质蛋白的转录/表达水平与纤毛虫形态特征存在相关性. 同时, 围绕纤毛虫系统进化中地位独特的弹跳虫, 新近完成了对其基因组的测序, 拼接与信息分析工作已在进展中. 此外, 本研究还包括对线粒体基因组所开展的工作: 围绕腹毛类的重要代表, 游仆虫(*Euplotes*)的线粒体基因组测序已初步完成.

下一步的工作将集中在已测序种类的后期分析上. 可以预期, 这项计划的开展, 将为解答纤毛虫系统发育历程、门内各大类群间的亲缘关系确定、追溯适应特征的起源、探讨遗传信息与表型间的联系、功能基因的进化等问题给出有价值的信息, 并将对基因树冲突、基因树-物种树冲突等问题的解答提供新

的依据.

4 纤毛虫表观遗传学研究

表观遗传学主要探讨非编码RNA、组蛋白翻译后修饰和DNA甲基化等表观遗传信息(作为DNA序列之外的信息储存)对生物体各项活动的调控机制^[34]. 自20世纪80年代起, 该领域已成为生物学研究最重要的热点之一. 纤毛虫因具有独特的二元核型及完善的组学信息而成为开展表观遗传学研究的优秀模式生物, 其中, 以四膜虫为材料有关组蛋白乙酰转移酶的刻画, 被誉为表观遗传学的奠基性工作^[35]; 扫描RNA(scRNA)模型的建立则被*Science*评为2002年的“年度重大科学发现”^[36].

图3 (网络版彩色)模式纤毛虫的基因组与转录组学. (a) 5种纤毛虫的BLAST相似性对比; (b) 根据基因组和转录组的同源基因数据库构建系统发育树; (c) 具沟急游虫及其他纤毛虫同源基因的比较; (d) 表层胞质蛋白基因家族在不同类群的转录本差异比较; (e) 在不同酶切浓度下通过MNase-seq获得的嗜热四膜虫单核小体DNA片段长度分布图

Figure 3 (Color online) Genomics and transcriptomics based on the model ciliates. (a) BLAST identity distribution between proteins of *Strombidium sulcatum* and that of *Oxytricha trifallax*, *Tetrahymena thermophila*, *Ichthyophthirius multifiliis* and *Paramecium tetraurelia*; (b) phylogenetic network inferred from the concatenated orthologs alignment using the neighbornet algorithm and the uncorrected distances; (c) comparisons of number of orthologs among *S. sulcatum* and other ciliates; (d) proportion of transcripts of the epiplasmic gene family in 13 ciliates divided into groups; (e) fragment size distribution of *T. thermophila* mono-nucleosomal DNA produced by MNase-seq with different MNase digestion concentration

现有工作表明, 纤毛虫在表观遗传调控方面具有极大的多样性: 尖毛虫中存在由长的非编码RNA指导的大规模的基因乱序及重排, 在四膜虫和草履虫中却几乎不存在^[37]. 即使是同属于寡膜类的草履虫和四膜虫也存在显著差异. 例如, 草履虫的基因区含有短的内部删除序列(internal eliminated sequence, IES), 需精确删除以避免移码突变^[38], 而四膜虫的IES几乎都位于非基因区, 大多被非精确删除^[39]. 以

四膜虫为模式的工作一直是国际间表观遗传学领域的重要分支, 目前主要涉及组蛋白翻译后修饰、非编码RNA以及端粒的研究, 热点问题包括: 扫描RNA模型的优化、常规复制及核内复制的表观遗传调控、营养生殖时期异染色体形成及转录调控以及接合生殖特异性基因在营养生殖时期的转录抑制等.

本团队新近的工作涉及以下两个焦点: DNA复制的表观遗传调控机制和RNA干扰通路和多梳蛋白抑制

通路的协同机制。针对第一个问题,发现了特异性修饰组蛋白H3第27位赖氨酸的单甲基化酶TXR1,并证实TXR1的敲除将严重影响DNA复制延伸的速度和保守度且导致单链DNA在复制起点附近大量积累,从而首次证明表观遗传信息的微扰可直接影响遗传信息的复制,为研究二者的协同复制奠定了基础(图4)^[40]。围绕第二个问题,初步刻画了与异染色体形成相关的四膜虫多梳蛋白复合体并针对RNA干扰通路的重要参与组分——解旋酶EMA1开展了深入的分析,证实了其RNA干扰通路和多梳蛋白抑制通路的介导因子^[36,41]。

人们今天已经对表观遗传调控机制有了很多了

解,但仍有众多问题需要解答,如纷繁的组蛋白修饰之间是如何协同作用的?这些修饰是如何精确地实现隔代传递的?今后的工作中,将特别关注组蛋白修饰对基因复制和表达的调控机制,包括以TXR1单甲基化酶为代表的表观遗传因子如何调控DNA复制延伸?小核中大量转座子如何沉默与激活?营养期的异染色体是如何形成及调控转录?相信这些工作的开展将了解表观遗传调控提供更广的视角和新的信息。

5 纤毛虫两型核间的基因组进化

纤毛虫最重要的特征之一是在同一个细胞内存

图 4 (网络版彩色)四膜虫的表观遗传学研究。(a) 组蛋白单甲基化酶TXR1的敲除导致的单链DNA的大量积累;(b) BrdU阳性细胞在不同细胞周期的分布情况;(c) $\Delta TXR1$ 细胞rDNA上富集更多的单链结合蛋白RPA1

Figure 4 (Color online) Epigenetic studies in *Tetrahymena*. (a) Deletion of H3K27 mono-methyltransferase TXR1 caused accumulation of single-strand DNA (ssDNA), as revealed by non-denaturing BrdU pulse-chase; (b) quantification of BrdU-positive TXR1 cells in different cell stages; (c) association of ssDNA-binding protein RPA1 with the rDNA chromosome, demonstrated by ChIP-QPCR. Note that more RPA1 are associated with in $\Delta TXR1$ cells

在两种不同类型的细胞核：大核(macronucleus, MAC)和小核(micronucleus, MIC). 前者也叫营养核, 负责营养生长期的基因表达, 除核残迹类, 纤毛虫的大核为多倍体; 小核也叫生殖核, 为二倍体, 在营养时期其染色体处于异染色质化, 没有基因表达, 其作用是将遗传信息传递给子代细胞^[42]. 正是因为这种核双态性, 纤毛虫的有性生殖具有十分独特的过程,

即接合生殖: 小核经过减数分裂由二倍体核变为单倍体核, 并与接合对中互配的细胞进行动核交换, 形成新的二倍体合子核, 最终发育为新的大、小核, 同时旧的大核解体. 在新大核形成过程中, 基因组会经历一次大范围的染色体重组过程, 包括序列的删除、染色体片段化、乱序基因的重排、基因多倍化等^[42-46]. 研究显示, 这一过程由RNA介导的表观遗传学控制^[36,47].

图5 (网络版彩色) 纤毛虫两型核间的基因组重组. (a) 纤毛虫生活史过程中的细胞核事件; (b) 选择性拼接可能是基因乱序进化过程的一个中间步骤, 并可能在物种形成中发挥作用; (c) 钩刺斜管虫——包含蛋白激酶域的蛋白基因家族成员由小核染色体到大核染色体可能的两个重组机制; (d) 大小核基因组拼接模式图; (e) 小核基因乱序的形成模型; (f) 钩刺斜管虫组氨酸酸性磷酸酶蛋白基因家族成员的大、小核序列对应图

Figure 5 (Color online) Genome rearrangements in ciliates. (a) Nuclear events in a generalized ciliate life cycles; (b) genome architecture drives evolution in ciliates, resulting in gene scrambling and perhaps even speciation; (c) two models showing possible arrangements for a protein kinase domain containing protein in the germline micronucleus in *Chilodonella uncinata*; (d) the three general patterns of genome rearrangements; (e) model for the origins of scrambled micronuclear genes; (f) schematic maps of the somatic and corresponding germline sequences of histidine acid phosphatase protein gene families in *Chilodonella uncinata*

由于纤毛虫高度的多样性,该重组程度和模式也因此呈现出显著的差异。例如,在嗜热四膜虫中,合子核中5条染色体会发生轻度的片段化和DNA删除,产生200条大小在100~1500 kb的大核染色体。每个大核染色体包含几百个基因,平均拷贝数为45^[42,43]。而一些高等纤毛虫如尖毛虫(*Oxytricha trifallax*),其大核染色体则高度片段化,合子核内的大约120条染色体经历片段化、乱序基因重排及选择性拼接后形成约24000条大核染色体,每条均小于15 kb,几乎每个基因均代表了一条染色体,且每条染色体约复制950~15000个拷贝^[29,42]。

前人在对管口类纤毛虫,钩刺斜管虫的研究中还发现一种特殊的染色体拼接模式——选择性拼接,即小核中的同一段序列会选择性地用来形成不同的大核染色体^[48]。为阐明基因组重组模式,利用钩刺斜管虫转录组分析及传统实验方法,对其基因乱序和不同基因家族的选择性拼接模式等进行了深入探讨。结果发现,借助高通量转录组数据分析,从中筛选出了112个候选的拼接基因家族,显示选择性拼接在其基因家族中广泛存在且这些基因很可能系该种所特有^[48]。进一步测定钩刺斜管虫中3个选择性拼接基因家族,发现选择性拼接比人们之前认为的更加复杂,且其不同基因家族具有不同的选择性拼接模式。根据已有数据提出了一个理论模型,即认为基因乱序由基因复制及之后的分化来演绎,而选择性拼接可能是基因乱序进化过程的中间形式,同时推测这种拼接极可能在纤毛虫物种形成中具有重要的作

用。该假说如得以证实,将为阐释基因组结构对基因家族进化的影响形成一个新观点,也为了解纤毛虫物种形成、分化机理提供一个新的视角(图5)^[49,50]。

此外,为深入了解基因组重组模式更为复杂和特殊的高等类群纤毛虫,新近扩增获得了两种海洋纤毛虫(*Pseudokeronopsis rubra*和*Uroleptopsis citrina*)的大核及小核Actin基因。分析结果初步显示,内部删除序列的获得及丢失频繁发生于腹毛亚纲纤毛虫中,而且片段化水平与基因乱序之间没有必然的联系。乱序的小核Actin基因仅存在于部分类群,而非普遍存在的现象说明小核Actin基因在类群间很可能是独立进化的^[51]。结合传统PCR和高通量测序的方法,最近以*Uroleptopsis citrina*为材料对其大核基因组进行了初步分析,同时,借助步移PCR技术,成功获取了该种ABC转运蛋白基因家族的3条大核染色体及对应的小核序列,结果显示,与尖毛虫的该基因相比,其基因片段化更为严重。这个现象的发现为深入了解纤毛虫独特的基因组重组现象提供了新的素材。

迄今为止,围绕纤毛虫大小核间基因组重组模式的研究尚处于探索阶段。人们的有关认知基本限于对四膜虫、草履虫及“一个基因一条染色体”模式的部分极端种类的研究^[42,48],而对于具有更高多样性和复杂性的大量海洋类群,有关其重组模式及机制仍一无所知。鉴于此,未来的工作将围绕不同代表类群的重组模式的特征、重组机制等问题开展研究,包括结合组学信息探讨基因组重组模式的进化关系,为揭示真核生物的物种分化与形成机理打开一扇窗口。

参考文献

- Greider C W, Blackburn E H. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell*, 1985, 43: 405–413
- Song W B, Warren A, Hu X Z. Free-living Ciliates in the Bohai and Yellow Seas, China. Beijing: Science Press, 2009 [宋微波, Warren A, 胡晓钟. 中国黄渤海的自由生纤毛虫. 北京: 科学出版社, 2009]
- Corliss J O. The Ciliated Protozoa: Characterization, Classification and Guide to the Literature Pergamon. New York: Oxford Press, 1979
- Gentekaki E, Kolisko M, Boscaro V, et al. Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. *Mol Phylogenet Evol*, 2014, 78: 36–42
- Lynn D H. The Ciliated Protozoa: Characterization, Classification and Guide to the Literature. Dordrecht: Springer Verlag, 2008
- Gao F, Katz L A. Phylogenomic analyses support the bifurcation of ciliates into two major clades that differ in properties of nuclear division. *Mol Phylogenet Evol*, 2014, 70: 240–243
- Gao F, Katz L A, Song W. Multigene-based analyses on evolutionary phylogeny of two controversial ciliate orders: Pleuronematida and Loxocephalida (Protista, Ciliophora, Oligohymenophorea). *Mol Phylogenet Evol*, 2013, 68: 55–63
- Huang J, Chen Z, Song W, et al. Three-gene based phylogeny of the Urostyloidea (Protista, Ciliophora, Hypotricha), with notes on classification of some core taxa. *Mol Phylogenet Evol*, 2014, 70: 337–347

- 9 Zhang Q, Simpson A, Song W. Insights into the phylogeny of systematically controversial haptorian ciliates (Ciliophora, Litostomatea) based on multigene analyses. *Proc Biol Sci*, 2012, 279: 2625–2635
- 10 Miao M, Warren A, Song W, et al. Analysis of the internal transcribed spacer 2 (ITS2) region of scuticociliates and related taxa (Ciliophora, Oligohymenophorea) to infer their evolution and phylogeny. *Protist*, 2008, 159: 519–533
- 11 Wang P, Gao F, Huang J, et al. A case study to estimate the applicability of secondary structures of SSU-rRNA gene in taxonomy and phylogenetic analyses of ciliates. *Zool Scr*, 2015, 44: 574–585
- 12 Yi Z, Dunthorn M, Song W, et al. Increasing taxon sampling using both unidentified environmental sequences and identified cultures improves phylogenetic inference in the Prorodontida (Ciliophora, Prostomatea). *Mol Phylogenet Evol*, 2010, 57: 937–941
- 13 Yi Z, Strüder-Kypke M, Hu X, et al. Sampling strategies for improving tree accuracy and phylogenetic analyses: A case study in ciliate protists, with notes on the genus *Paramecium*. *Mol Phylogenet Evol*, 2014, 71: 142–148
- 14 de Vargas C, Audic S, Henry N, et al. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 2015, 348: 1261605
- 15 Pawlowski J, Audic S, Adl S, et al. CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol*, 2012, 10: e1001419
- 16 Hebert P D, Cywinska A, Ball S L, et al. Biological identifications through DNA barcodes. *Proc Biol Sci*, 2003, 270: 313–321
- 17 Barth D, Przyboś E, Fokin S I, et al. Cytochrome b sequence data suggest rapid speciation within the *Paramecium aurelia* species complex. *Mol Phylogenet Evol*, 2008, 49: 669–673
- 18 Lynn D H, Strüder-kypke M C. Species of *Tetrahymena* identical by small subunit rRNA gene sequences are discriminated by mitochondrial cytochrome c oxidase I gene sequences. *J Eukaryot Microbiol*, 2006, 53: 385–387
- 19 Chantangsi C, Lynn D H, Brandl M T, et al. Barcoding ciliates: A comprehensive study of 75 isolates of the genus *Tetrahymena*. *Int J Syst Evol Microbiol*, 2007, 57: 2412–2425
- 20 Stoeck T, Przybos E, Dunthorn M. The D1-D2 region of the large subunit ribosomal DNA as barcode for ciliates. *Mol Ecol Resour*, 2014, 14: 458–468
- 21 Santoferrara L F, McManus G B, Alder V A. Utility of genetic markers and morphology for species discrimination within the order Tintinnida (Ciliophora, Spirotrichea). *Protist*, 2013, 164: 24–36
- 22 Strüder-Kypke M C, Lynn D H. Comparative analysis of the mitochondrial cytochrome c oxidase subunit I (COI) gene in ciliates (Alveolata, Ciliophora) and evaluation of its suitability as a biodiversity marker. *Syst Biodivers*, 2010, 8: 131–148
- 23 Zhao Y, Yi Z, Gentekaki E, et al. Utility of combining morphological characters, nuclear and mitochondrial genes: An attempt to resolve the conflicts of species identification for ciliated protists. *Mol Phylogenet Evol*, 2016, 94: 718–729
- 24 Zhao Y, Gentekaki E, Yi Z, et al. Genetic differentiation of the mitochondrial cytochrome oxidase c subunit I gene in genus *Paramecium* (Protista, Ciliophora). *PLoS One*, 2013, 8: e77044
- 25 Song W B, Xu K D, Shi X L, et al. Progress in Protozoology. Qingdao: Qingdao Ocean University Press, 1999 [宋微波, 徐奎栋, 施心路, 等. 原生动物学专论. 青岛: 青岛海洋大学出版社, 1999]
- 26 Dunthorn M, Klier J, Bunge J, et al. Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity. *J Eukaryot Microbiol*, 2012, 59: 185–187
- 27 Kodama Y, Suzuki H, Dohra H, et al. Comparison of gene expression of *Paramecium bursaria* with and without *Chlorella variabilis* symbionts. *BMC Genomics*, 2014, 15: 183
- 28 McGrath C L, Gout J F, Doak T G, et al. Insights into three whole-genome duplications gleaned from the *Paramecium caudatum* genome sequence. *Genetics*, 2014, 197: 1417–1428
- 29 Swart E C, Bracht J R, Magrini V, et al. The *Oxytricha trifallax* macronuclear genome: A complex eukaryotic genome with 16000 tiny chromosomes. *PLoS Biol*, 2013, 11: e1001473
- 30 Eisen J A, Coyne R S, Wu M, et al. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol*, 2006, 4: e286
- 31 Cervantes M D, Hamilton E P, Xiong J, et al. Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. *PLoS Biol*, 2013, 11: e1001518
- 32 Chen X, Zhao X, Liu X, et al. Phylogenomics of non-model ciliates based on transcriptomic analyses. *Protein Cell*, 2015, 6: 373–385
- 33 Feng J M, Jiang C Q, Warren A, et al. Phylogenomic analyses reveal subclass Scuticociliatia as the sister group of subclass Hymenostomatia within class Oligohymenophorea. *Mol Phylogenet Evol*, 2015, 90: 104–111
- 34 Allis C D, Jenuwein T, Reinberg D. Overview and concepts. In: Allis C D, Jenuwein T, Reinberg D, et al, eds. *Epigenetics*. New York: Cold Spring Harbor Laboratory Press, 2007. 23–62
- 35 Brownell J E, Allis C D. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in *Tetrahymena* macronuclei. *Proc Natl Acad Sci USA*, 1995, 92: 6364–6368

- 36 Mochizuki K, Fine N A, Fujisawa T, et al. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell*, 2002, 110: 689–699
- 37 Chen X, Bracht J R, Goldman A D, et al. The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. *Cell*, 2014, 158: 1187–1198
- 38 Arnaiz O, Mathy N, Baudry C, et al. The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: Evolutionary dynamics of internal eliminated sequences. *PLoS Genet*, 2012, 8: e1002984
- 39 Yao M C, Fuller P, Xi X. Programmed DNA deletion as an RNA-guided system of genome defense. *Science*, 2003, 300: 1581–1584
- 40 Gao S, Xiong J, Zhang C, et al. Impaired replication elongation in *Tetrahymena* mutants deficient in histone H3 Lys 27 monomethylation. *Genes Dev*, 2013, 27: 1662–1679
- 41 Aronica L, Bednenko J, Noto T, et al. Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev*, 2008, 22: 2228–2241
- 42 Prescott D M. The DNA of ciliated protozoa. *Microbiol Rev*, 1994, 58: 233–267
- 43 Chalker D L, Yao M C. DNA elimination in ciliates: Transposon domestication and genome surveillance. *Annu Rev Genet*, 2011, 45: 227–246
- 44 Goldman A D, Landweber L F. *Oxytricha* as a modern analog of ancient genome evolution. *Trends Genet*, 2012, 28: 382–388
- 45 Nowacki M, Shetty K, Landweber L F. RNA-mediated epigenetic programming of genome rearrangements. *Annu Rev Genomics Hum Genet*, 2011, 12: 367–389
- 46 Riley J L, Katz L A. Widespread distribution of extensive chromosomal fragmentation in ciliates. *Mol Biol Evol*, 2001, 18: 1372–1377
- 47 Angeleska A, Jonoska N, Saito M, et al. RNA-guided DNA assembly. *J Theor Biol*, 2007, 248: 706–720
- 48 Katz L A, Kovner A M. Alternative processing of scrambled genes generates protein diversity in the ciliate *Chilodonella uncinata*. *J Exp Zool B Mol Dev Evol*, 2010, 314: 480–488
- 49 Gao F, Roy S W, Katz L A. Analyses of alternatively processed genes in ciliates provide insights into the origins of scrambled genomes and may provide a mechanism for speciation. *MBio*, 2015, 6: e01998–e01914
- 50 Gao F, Song W, Katz L A. Genome structure drives patterns of gene family evolution in ciliates, a case study using *Chilodonella uncinata* (Protista, Ciliophora, Phyllopharyngea). *Evolution*, 2014, 68: 2287–2295
- 51 Chen T, Yi Z, Huang J, et al. Evolution of the germline actin gene in hypotrichous ciliates: Multiple nonscrambled IESs at extremely conserved locations in two urostylids. *J Eukaryot Microbiol*, 2015, 62: 188–195
- 52 Zheng W B, Gao F. Genome rearrangements between the dimorphic nuclei in ciliates (in Chinese). *Chin Bull Life Sci*, 2015, 27: 486–494 [郑维波, 高凤. 纤毛虫两型核之间的基因组重组. *生命科学*, 2015, 27: 486–494]

On molecular biology of ciliated protozoa: Frontier areas and progresses

YI ZhenZhen¹, MIAO Miao², GAO Shan³, GAO Feng³ & SONG WeiBo³

¹ Guangzhou Key Laboratory of Subtropical Biodiversity and Biomonitoring, South China Normal University, Guangzhou 510631, China;

² Savaid Medical School, University of Chinese Academy of Sciences, Beijing 100049, China;

³ Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao 266003, China

Ciliated protists are important organisms which play significant roles in studies on cell biology, genetics, biodiversity, biogeography and ecology, environmental biology etc. In last decade, molecular investigations on phylogeny, barcoding, epigenetics, genomics, and gene evolution have been performed by the authors' groups and the main results can be summarized as follows: (i) Phylogenetic relationships within over 50 classes/orders are analyzed, and two new classes, six new orders/suborders are established by the authors' groups, based on more than 2000 new sequences. Additionally, theoretical researches focusing on gene marker selection and topological optimization are also performed, in order to provide theoretical guidance for future study. We suggested that unidentified environmental sequences, and secondary structures of rRNA gene could be used for inferring phylogenetic relationships of ciliates, at least on some taxon level. And we also discussed the impact of missing data present in concatenated analyses, and paralogs of protein-coding genes on phylogenetic analyses. (ii) Transcriptome/genome of six representative ciliates (*Strombidium sulcatum*, *Uronema marinum*, *Metanophrys similis*, *Euplotes vannus*, *Uroleptopsis citrina*, *Paramecium caudatum*) were sequenced and analyzed, which will help to find out the evolutionary position of key species, the relationship between expression of the functional gene family and morphological characteristics, and to understand the association between genetic information and phenotype; Also, we compared mitochondrial genome of the genus *Euplotes* species, thus found the gene order and transcriptional directions were the same among three species, but 3 split protein coding genes (*nad9*, *nad2* and *rpl16*) were predicted only in mt-genome of *E. vannus*. (iii) Knockout of histone methyltransferase TXR1 (*Tetrahymena* Trithorax related 1), specific for H3 lysine 27 mono-methylation (H3K27me1), leads to severe replication stress in the model organism *Tetrahymena thermophila*. Consequently, cells mobilize the DNA damage response, indicating that perturbation of epigenetic information can directly affect replication elongation. (iv) Alternative processing is extensive among gene families in *Chilodonella uncinata* which are likely to be *C. uncinata* specific; and the processing patterns are different among gene families which is more complex than previously believed. We propose a new model whereby gene scrambling evolves by gene duplication followed by partial and reciprocal degradation of the duplicate sequences. In this model, alternative processing represents an intermediate step in the evolution of scrambled genes and yields important insights into the mechanism for speciation in ciliates. (v) Four gene markers, viz. *COI*, *ITS1*, *ITS2*, *D2* region of 28S rRNA gene, are evaluated as promising candidate barcoding markers, by comparing 8 gene makers on the same organisms for the first time. DNA barcoding sequences of 83 species covering subclass Peniculia, Oligotrichia and Euplotia are firstly sequenced. And sampling information, morphological characters and pictures of these species are also recorded for constructing DNA barcoding database of ciliates.

ciliated protozoa, phylogeny, barcoding, epigenetics, genome, evolution

doi: 10.1360/N972016-00064