

文章编号: 1000-5463(2013)05-0001-06

原生生物基因重复研究进展

黄丽娟, 伊珍珍, 林晓凤*

(华南师范大学生命科学学院, 广东省高等学校生态与环境科学重点实验室, 广东广州 510631)

摘要:详述了原生生物的大规模和小规模基因重复2种类型,以及重复基因新功能化、亚功能化和假基因化等3种分化途径的研究现状,简要介绍了重复基因保留偏好性和3种进化模型等方面的相关研究,揭示了基因重复对原生生物进化的重要意义.建议多开展原生生物非模式生物的研究工作.

关键词:基因重复;原生生物;基因保留;进化模型

中图分类号:Q-1 **文献标志码:**A doi:10.6054/j.jscnun.2013.07.001

基因重复(Gene Duplication)指含有基因的DNA片段发生重复,产生1个与原基因相似的基因或碱基序列.由基因重复产生的基因称为重复基因,或旁系同源基因(Paralogs)^[1],通过不等交换、逆转录转座或全基因组重复等途径产生^[2].根据重复区域的大小,基因重复可分为2种类型:(1)小规模基因重复,主要是单个基因重复;(2)大规模基因重复,包括部分基因组重复以及整个基因组重复(多倍体化)^[3].单个基因和部分基因组的重复主要通过不等交换产生,而全基因组重复是有丝分裂或减数分裂过程中的错误引发^[4].

在原核生物和真核生物中,都存在大量的重复基因现象^[5],它为基因家族的产生提供了最原始的遗传物质基础^[6].例如,在原核生物肺炎支原体(*Mycoplasma pneumoniae*)的基因组中,共生同源基因家族(Paralogous Gene Family)占全基因组的44%^[7].在动物中,秀丽隐杆线虫(*Caenorhabditis elegans*)和黑腹果蝇(*Drosophila melanogaster*)的重复基因分别占全基因组的49%和41%^[5].植物基因重复现象更为普遍,大约70%~80%的被子植物物种存在基因重复或多倍化过程,其中拟南芥(*Arabidopsis thaliana*(L.) Heynh.)至少经历了3次全基因组重复^[8].单细胞的真核生物中,也存在重复基因,特别是具有双型核的纤毛门,大核存在大量重复的染色体片段,某些物种的大核染色体片段,重复次数达950到15 000次^[9].

基因重复导致同一个基因组内存在2个或者2个以上拷贝的同源基因序列,这可能造成功能上的

冗余^[5],并受到剂量效应的调节^[10].它与生物体基因组大小的进化、新基因的起源以及物种的分化等都密切相关^[4-5],通过对基因重复的研究,能更好地理解基因家族的形成、基因表达水平的多样性.目前对于高等真核生物基因重复的现象报道及进化机制研究日趋增加^[8],但对于单细胞真核生物基因重复的研究还处于初期阶段.相关研究证实,单细胞和多细胞真核生物通过基因重复产生的基因家族大小与选择压力的关系有所不同^[2].本文选取原生生物作为单细胞真核生物的代表,总结近年来原生生物重复基因的研究进展,包括基因重复类型、保留偏好性、分化途径,并讨论重复基因的3种进化模型,揭示基因重复对生物进化的重要性,提出关于后续研究开展方向的建议.

1 基因重复的2种类型

1.1 大规模基因重复

模式生物草履虫 *Paramecium tetraurelia* 至少经历了3次连续的全基因组重复事件,共产生了将近40 000个编码蛋白的基因^[11].依照基因重复发生的前后顺序,可以分为早期、中期和近期全基因组重复,这3个时期产生的旁系同源基因分别占8%、24%和51%^[11].在囊泡藻 *Blastocystis* sp. 的基因组中,DENOEUDE等^[12]发现了占全基因组39%左右的重复区域,并推测这是全基因组重复的结果.SUN等^[13]在鞭毛虫 *Giardia lamblia* 基因组中发现了2 403个重复基因,大约占全基因组的40%,进化距

离分析表明 *G. lamblia* 基因组在进化过程中发生了 2 次大规模的基因重复事件. LORENZI 等^[14] 发现阿米巴虫 *Entamoeba histolytica* 基因组存在部分基因组重复, 并且约 20% 的基因组包含转座子, 推测部分基因组重复与某些基因家族含有转座子密切相关.

1.2 小规模基因重复

目前, 对原生生物小规模基因重复的研究主要涉及到基因家族, 如肌动蛋白 (Actin) 基因家族和微管蛋白 (Tubulin) 基因家族, 以及高度重复的核糖体 RNA 基因 (rDNA) 等^[15-17].

1.2.1 肌动蛋白基因 肌动蛋白是真核细胞中普遍存在的一种蛋白, 在细胞运动、细胞分裂和胞内运输等方面发挥重要作用^[15]. 肌动蛋白基因是一类多基因家族, 经过基因重复在生物中存在许多旁系同源基因, 并且经过旁系同源基因分化, 在种间或种内形成多种基因亚型. 草履虫 *Paramecium tetraurelia* 存在 9 种肌动蛋白基因亚型, 并且在细胞中有不同的定位与功能^[18]. KIM 等^[15] 分析了纤毛门多种类群的肌动蛋白基因序列, 发现纤毛门肌动蛋白旁系同源基因高度分化. 表壳目 (Arcellinida) 阿米巴虫的 3 个独立遗传谱系 (1 个 *Arcella hemisphaerica*、2 个 *A. vulgaris*) 中, 每一个谱系的肌动蛋白基因包含 40~50 个旁系同源基因, 其中没有任何一个谱系与其他谱系的旁系同源基因相同, 基因分化程度达到了 29%^[19]. 目前本课题组选取伪角毛虫属 (*Pseudokeronopsis*) 的 4 个形态相似种 (红色伪角毛虫、黄色伪角毛虫、肉色伪角毛虫、拟红色伪角毛虫) 的 8 个种群作为实验材料, 研究大核肌动蛋白基因重复现象, 结果显示伪角毛虫属中肌动蛋白基因至少存在 2 种基因亚型 (未发表资料).

1.2.2 微管蛋白基因 微管蛋白是真核细胞的骨架蛋白之一, 在保持细胞形状、细胞运动、胞内物质运输等方面起着重要作用^[20]. 微管蛋白基因超家族共有 9 个基因成员 (α 、 β 、 γ 、 δ 、 ϵ 、 η 、 θ 、 ζ 和 κ -微管蛋白基因), 在草履虫 *Paramecium tetraurelia* 的单倍体小核基因组中 9 种微管蛋白基因均存在^[17], 其中 α -微管蛋白基因包含 2 个旁系同源基因, 而 β -微管蛋白基因包含 3 个旁系同源基因. α -微管蛋白基因家族中的 2 个旁系同源基因的序列非常相似, 意味着它们是由最近的一个共同祖先通过基因重复形成的, 或者通过基因转换而导致同源化^[17]. DELGADO 等^[20] 利用分子杂交技术, 证实纤毛虫 *Histriculus cavicola*, *Euplotes euryostomus* 和 *Stentor coeruleus* 双倍体小核基因组中 α -微管蛋白旁系同源基因数目分别为 3、2 和 9 以上. 纤毛虫大核发育过程中, 染色体高

度片段化并经历多次基因重复, 因此一种微管蛋白基因在大核中存在多份拷贝, 如斜管虫 *Chilodonella uncinata*, 大核中至少存在 4 种 β -微管蛋白旁系同源基因^[21].

1.2.3 核糖体 RNA 基因 核糖体 RNA 基因 (rDNA) 包括小亚基核糖体 RNA 基因 (SSU rDNA)、大亚基核糖体 RNA 基因 (LSU rDNA) 和 5.8S rDNA. rDNA, 属于高度重复序列. 硅藻 (Diatoms) 单个细胞中 rDNA 拷贝数在 61~36 896 范围内^[22], 双鞭毛藻 (Dinoflagellates) 单个细胞中 rDNA 拷贝数大致为 200~1 200^[23], 微藻 (Microalgae) 单个细胞中 rDNA 拷贝数约为 1~12 000^[24]. 纤毛虫大核核糖体 RNA 基因则具有更高的拷贝数, 如棘尾虫 *Stylonychia lemnae* 单个细胞中 rDNA 拷贝数多达 400 000, 钟形虫 *Vorticella* sp. 单个细胞中 rDNA 拷贝数为 315 786, 尖毛虫 *Oxytricha nova* 单个细胞中 rDNA 拷贝数约为 200 000, 拟尾虫 *Pseudotontonia* sp. 单个细胞中 rDNA 拷贝数为 172 889, 拟铃虫 *Tintinnopsis* sp. 单个细胞中 rDNA 有 126 372 个拷贝^[17]. JOHN-SON 等^[25] 发现侧口纲 (Litostomatea) 的 2 种纤毛虫 *Myrionecta rubra* 和 *Mesodinium pulex* 的 SSU rDNA 与其它侧口纲纤毛虫高度分化, 推测由基因重复引起.

1.2.4 其他基因 组蛋白 (Histones) 是真核生物体细胞染色质中的碱性蛋白质, 有 H2A、H2B、H3、H4 和 H1 共 5 种基因类型, 目前为止在原生动物中只发现了前 4 种, 并且 H2A、H3 基因家族含有大量变异度高的旁系同源基因^[26]. 例如, 毛滴虫 *Trichomonas vaginalis* 中分别含有 17、14、21、22 个拷贝的 H2A、H2B、H3、H4 基因^[26].

UDA 等^[27] 在模式生物四膜虫 *Tetrahymena* 基因组中发现了 2 种类型的精氨酸激酶 (Arginine Kinase), 一种只含有 40×10^3 的亚基; 另一种含有 2 个结构域, 是经过基因重复及融合事件产生的, 并且这 2 个事件至少独立地发生了 4 次.

热激蛋白 Hsp70 是一类多基因家族, 包含大量分化程度高的旁系同源基因, BUDIN 等^[28] 从草履虫 *Paramecium tetraurelia* 和游仆虫 *Euplotes aediculatus* 中分别克隆出 8 种 Hsp70 旁系同源基因.

2 重复基因保留偏好性

基因重复发生后, 特别是当整个基因及调控区域都发生重复的时候, 常造成 2 个旁系同源基因功能上的冗余. 由于自然选择的作用, 通常只有一个旁系同源基因能够被保留下来, 另一个拷贝发生突变

导致功能丧失而成为假基因,并逐渐消亡^[29].相关研究表明,重复基因的保留与丢失并不是一个完全随机的过程,而是存在一定的偏好性^[30-31].

目前研究得较为清楚的草履虫 *Paramecium tetraurelia* 的基因组中,属于相同代谢途径或蛋白复合体的基因有共同的基因丢失模式,并且表达量高的基因(如信号分子和转录因子的基因)在重复基因中占有较高的比例. AURY 等^[11]的研究表明,全基因组重复后许多基因的保留与剂量效应有关.在全基因组重复后的短时期内,属于同一个调控网络或代谢途径的相互作用蛋白之间要保持相对剂量而被保留下来,这种模型被称为剂量效应模型,但这种模型可能只适用于基因组重复发生的较短时间内,给重复基因的新功能化提供机会^[32],对于时间久远的重复基因并不适用. EDGER 等^[10]进一步提出转录因子基因等剂量敏感的基因在全基因组重复后会大量保留,但在小规模基因重复后,却有较低的保留率.与代谢相关的基因在全基因组重复后会比其他基因保留率高,但如果代谢基因在全基因组重复发生之前有很高的拷贝数,其保留率将会大大降低^[33].

3 重复基因的分化

基因重复作为一种新基因的产生机制,为生物新功能的产生提供了最原始的遗传物质基础^[4].经过自然选择的作用,保留下来的重复基因,大致面临以下3种不同的命运^[5,8]:(1)其中1个拷贝保留了原始的功能,另1个拷贝新功能化;(2)2个拷贝发生了亚功能化,分担了原基因的功能;(3)其中1个拷贝发生突变导致非功能化成为假基因.

3.1 新功能化

新功能化是指含有基因的DNA片段经重复和歧化后获得新的功能,经自然选择后被保留下来成为新的基因,这是重复基因能够被长期保留的机制^[5].生物界新功能化最经典的一个例子是,哺乳动物叶猴科(Colobine Monkey)的核糖核酸酶RNAse1基因通过基因重复,产生了2个基因,分别是RNAse1a基因和RNAse1b基因^[34].9个氨基酸的替换导致RNAse1b蛋白的最佳pH改变,使这些猴子可以将叶子作为首要食物来源,而不像其他猴子一样食用果实和昆虫^[34].原生动物草履虫 *Paramecium tetraurelia* 中发现了一种新的微管蛋白, η -微管蛋白,这种微管蛋白的基因与之前在草履虫发现的微管蛋白(α 、 β 、 γ 、 δ 和 ε -微管蛋白)基因序列相似

度很低,系统分析结果显示 η -微管蛋白基因与其他微管蛋白基因亲缘关系很远^[35].与在真核生物中保守存在并负责组装基体的 δ -微管蛋白不同, η -微管蛋白目前发现只存在于纤毛虫中,并参与纤毛基体的复制^[11]. RUIZ 等^[35]发现 η -微管蛋白基因突变,不仅抑制基体的复制,还导致 γ -微管蛋白与基体分离,推测 η -微管蛋白起着紧密连接 γ -微管蛋白与基体的作用.

3.2 亚功能化

亚功能化,即重复基因的不同拷贝分别执行原基因的部分功能,各基因拷贝之间的功能是互补的.例如,SEHRING 等^[16]在草履虫 *Paramecium tetraurelia* 中发现了9种肌动蛋白的重复基因,它们在细胞中的分布以及功能是有差别的,例如act4和act9与细胞分裂有关,而act2、act3、act4及act9都与纤毛虫的运动行为有关,不同的重复基因沉默可使纤毛虫呈现不同的游动方式.本课题组沿着该思路开展工作,伪角毛虫属(*Pseudokeronopsis*)肌动蛋白基因的各基因亚型在细胞中的定位与功能,将是本课题组近期研究的重要内容. RASTOGI 等^[36]提出基因亚功能化在基因重复后的初始阶段,对于重复基因的保留具有重要的意义,但亚功能化只是新功能化的一个过渡阶段,而不是重复基因的最终命运.

3.3 假基因化

假基因化是指突变导致基因结构破坏,使本来具有正常功能的基因非功能化^[5].假基因化是重复基因常见的一种分化途径,草履虫 *Paramecium tetraurelia* 经历了3次连续的全基因组重复后,在40000个编码蛋白的基因中存在近1500个假基因^[11]. PEACOCK 等^[37]在3种利什曼原虫 *Leishmania major*、*L. infantum* 和 *L. braziliensis* 的基因组中分别发现了97、41和161个假基因,并推测假基因的形成和基因丢失是导致这3个种基因组差异的主要因素.毛滴虫 *Trichomonas vaginalis* 的基因组大约有60000个基因, CUI 等^[38]发现假基因占了很大的比例,其中由于无义突变引起的假基因约有3000个,由于阅读框架移动产生的假基因数目与之相近或有更多.

4 重复基因的3种进化模型

重复基因的进化模型主要有3种:发散进化(Divergent evolution)、协同进化(Concerted evolution)、生与死进化(Birth-and-death evolution)模型^[39].

发散进化指的是经基因重复产生的旁系同源基因,随着获得新的基因功能而逐渐分化,例如编码肌

红蛋白及血红蛋白 α 、 β 、 γ 和 δ 链的基因^[39].

协同进化是指种内的旁系同源基因比种间相同基因家族的成员亲缘关系更近,即使基因重复事件发生在物种分化之前.此进化模型中旁系同源基因的高度相似主要通过不等交换或者同源重组导致基因转换(Gene Conversion)维持^[40],同源重组能有效地复制部分序列并覆盖在其它的同源区,这个模型能很好解释 rRNA 基因及毛滴虫 *Trichomonas vaginalis* 聚泛素(Polyubiquitin)基因的进化^[39, 41].

生与死进化模型阐述的是多基因家族内旁系同源基因的产生、分化及丢失的过程,与协同进化模型不同,此模型中的基因家族成员是独立进化的^[40].在这个模型中,基因重复产生新的基因,一部分重复基因可以在基因组内保留很长一段时间,而一部分基因由于有害突变导致功能丧失而成为假基因或者丢失^[39].组氨酸 H4 基因家族,及疟原虫 *Plasmodium* 的裂殖子表面蛋白(msp)-7 基因家族的进化都是符合生与死进化模型,在进化过程中经历了纯化选择^[42-43].

协同进化模型和生与死进化模型是存在争议的,因为当序列差别很小的时候,很难区分这2个模型^[39].并且,基因家族的进化是很复杂的,并不是完全符合某一个进化模型,例如双鞭毛藻(Dinoflagellates)的肌动蛋白基因家族和线虫(Nematodes)的热激蛋白 Hsp70 基因家族,在进化过程中都经历过协同进化及生与死进化历程^[44-45].

5 结论

基因重复是一种非常普遍的生物学过程,它最显著的贡献就是为生物的进化提供了最原始的遗传物质基础,通过突变和选择的作用,可以产生新基因或亚功能化的基因,促使物种的分化和多样性的增多^[46].AURY 等^[11]的研究表明,全基因组重复与物种的分化有关,近期全基因组重复导致草履虫 *Paramecium aurelia* 分化形成 15 个姐妹种.并且,与单拷贝基因的表达相比较,重复基因的不同拷贝可以在特定的时间和空间表达,促使基因表达水平多样性的提高^[2].

近年来原生生物重复基因的研究取得了很大的进展,但是许多关于重复基因的问题仍然没有得到解决.首先,目前完成基因组测序的原生生物比较少,例如纤毛虫只有 4 种完成了大核基因组测序,分别是草履虫 *Paramecium tetraurelia*、四膜虫 *Tetrahymena thermophila*、小瓜虫 *Ichthyophthirius multifiliis*

和尖毛虫 *Oxytricha trifallax*^[11, 47-49],除前面提及的草履虫外,四膜虫和小瓜虫基因组中也存在大量重复基因,但是由于基因重复在物种进化过程中发生的时间跨度大,没有证据表明是属于部分或整个基因组重复^[47-48].其次,没有清晰地了解基因家族(如肌动蛋白、微管蛋白基因家族)在原生生物中的旁系同源基因种类、拷贝数及各自的功能^[35].再次,亚功能化、剂量补偿等理论可解释重复基因的保留途径,但是目前没有清晰的机制解释重复基因如何逐步形成新功能^[33].最后,对于原生生物重复基因的自然选择途径、进化速率、进化模型等方面还没有明确的认识.

上述问题的存在,主要是因为目前已开展的原生生物基因重复研究基本局限于少数几种模式生物,如草履虫 *Paramecium tetraurelia*、四膜虫 *Tetrahymena thermophila* 和毛滴虫 *Trichomonas vaginalis* 等^[11, 14, 38].因此,后续研究应该充分利用快速发展的分子遗传学技术,对更多原生生物(尤其是非模式生物)的重复基因类型、产生、保留及进化机制等进行深入的研究.作为原生生物非模式生物的探索性工作,如 SUN 等^[13]报道寄生性鞭毛虫 *Giardia lamblia* 基因组发生了 2 次大规模基因重复事件,作者发现大部分的重复基因表达的是变化的特异表面蛋白(Variant-specific Surface Proteins),这种蛋白与寄生虫逃避宿主的免疫系统的攻击有关,表明基因重复对 *G. lamblia* 寄生生活方式的起源与进化起着重要的作用;DELGADO 等^[20]测定了纤毛虫 *Histriculus cavicola*, *Euplotes euryostomus* 和 *Stentor coeruleus* 双倍体小核基因组中 α -微管蛋白旁系同源基因数目,有助于进一步认识旋唇纲(Spirotrichea)纤毛虫中 α -微管蛋白基因家族的基因类型和数目.这也是本课题组目前在原生生物基因重复领域开展工作的研究思路之一,作者近期开展的基因重复研究中,选取非模式生物伪角毛虫属(*Pseudokeronopsis*)的几个相近种作为实验材料,探讨伪角毛虫属肌动蛋白基因的重复基因类型、拷贝数、各基因亚型在细胞中的定位与功能,以期对纤毛虫肌动蛋白重复基因的保留及功能分化有进一步的了解.

参考文献:

- [1] KOONIN E V. Orthologs, paralogs, and evolutionary genomics[J]. Annu Rev Genet, 2005, 39: 309-338.
- [2] 彭贵子,陈玲玲,田大成. 基因重复的研究进展[J]. 遗传, 2006, 28(7): 886-892.
- [3] HAKES L, PINNEY J W, LOVELL S C, et al. All duplicates are not equal: The difference between small-scale

- and genome duplication [J]. *Genome Biol*, 2007, 8(10):R209.
- [4] 李昕, 杨爽, 彭立新, 等. 新基因的起源与进化[J]. *科学通报*, 2004, 49(13):1219–1225.
- [5] ZHANG J Z. Evolution by gene duplication: An update [J]. *Trends Ecol Evol*, 2003, 18(6):292–298.
- [6] RODRIGUEZ-TRELLES F, TARRIO R, AYALA F J. Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the xanthine dehydroge nase gene[J]. *Proc Natl Acad Sci*, 2003, 100(23):13413–13417.
- [7] HIMMELREICH R, HILBERT H, PLAGENS H, et al. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae* [J]. *Nucleic Acids Res*, 1996, 24(22):4421–4429.
- [8] 孙红正, 葛颂. 重复基因的进化——回顾与进展[J]. *植物学报*, 2010, 45(1):13–22.
- [9] PRESCOTT D M. The DNA of ciliated protozoa[J]. *Microbiol Rev*, 1994, 50(2):233–267.
- [10] EDGER P P, PIRES J C. Gene and genome duplications: The impact of dosage-sensitivity on the fate of nuclear genes[J]. *Chromosome Res*, 2009, 17(5):699–717.
- [11] AURY J M, JAILLON O, DURET L, et al. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia* [J]. *Nature*, 2006, 444(7116):171–178.
- [12] DENOEU D F, ROUSSEL M, NOEL B. Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite[J]. *Genome Biol*, 2011, 12:R29.
- [13] SUN J, JIANG H F, FLORES R. Gene duplication in the genome of parasitic *Giardia lamblia*[J]. *BMC Evol Biol*, 2010, 10(49):1–8.
- [14] LORENZI H A, PUIU D, MILLER J R, et al. New assembly, reannotation and analysis of the *Entamoeba histolytica* genome reveal new genomic features and protein content information [J]. *PLoS Negl Trop Dis*, 2010, 4(6):e716.
- [15] KIM O T P, YURA K, GO N, et al. Highly divergent actins from karyorelictean, heterotrich, and litostome ciliates[J]. *J Eukaryot Microbiol*, 2004, 51(2):227–233.
- [16] LIBUSOVA L, DRABER P. Multiple tubulin forms in ciliated protozoan *Tetrahymena* and *Paramecium* species [J]. *Protoplasma*, 2006, 227(2/3/4):65–76.
- [17] GONG J, DONG J, LIU X, et al. Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of *Oligotrich* and *Peritrich* ciliates[J]. *Protist*, 2013, 164(3):369–379.
- [18] SEHRING I M, REINER C, MANSFELD J. A broad spectrum of actin paralogs in *Paramecium tetraurelia* cells display differential localization and function[J]. *J Cell Sci*, 2007, 120(1):177–190.
- [19] LAHR D J, NGUYEN T B, BARBERO E, et al. Evolution of the actin gene family in testate lobose amoebae (*Arcellinida*) is characterized by two distinct clades of paralogs and recent independent expansions[J]. *Mol Biol Evol*, 2011, 28(1):223–236.
- [20] DELGADO P, CALVO P, VISCOGLIOSI E. Estimation of the number of α -tubulin genes in three ciliated protozoa[J]. *Archiv für Protistenkunde*, 1995, 145(1/2):105–111.
- [21] ZUFALL R A, KATZ L A. Micronuclear and macronuclear forms of beta-tubulin genes in the ciliate *Chilodonella uncinata* reveal insights into genome processing and protein evolution [J]. *J Eukaryot Microbiol*, 2007, 54(3):275–282.
- [22] GODHE A, ASPLUND M, HARNSTROM K. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR[J]. *Appl Environ Microbiol*, 2008, 74(23):7174–7182.
- [23] GALLUZZI L, PENNA A, BERTOZZINI E. Development of a real-time PCR assay for rapid detection and quantification of *Alexandrium minutum* (a dinoflagellate) [J]. *Appl Environ Microbiol*, 2004, 70:1199–1206.
- [24] ZHU F, MASSANA R, NOT F, et al. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene[J]. *FEMS Microbiol Ecol*, 2005, 52(1):79–92.
- [25] JOHNSON M D, TENGS T, OLDACH D W, et al. Highly divergent SSU rRNA genes found in the marine ciliates *Myrionecta rubra* and *Mesodinium pulex* [J]. *Protist*, 2004, 155(3):347–359.
- [26] DALMASSO M C, SULLIVAN W J Jr, ANGEL S O. Canonical and variant histones of protozoan parasites [J]. *Front Biosci*, 2011, 16:2086–2105.
- [27] UDA K, FUJIMOTO N, AKIYAMA Y, et al. Evolution of the arginine kinase gene family [J]. *Comp Biochem Phys D*, 2006, 1(2):209–218.
- [28] BUDIN K, PHILIPPE H. New insights into the phylogeny of eukaryotes based on ciliate Hsp70 sequences[J]. *Mol Biol Evol*, 1998, 15(8):943–956.
- [29] QIAN W F, LIAO B Y, CHANG A Y F, et al. Maintenance of duplicate genes and their functional redundancy by reduced expression [J]. *Trends Genet*, 2010, 26(10):425–430.
- [30] ADAMS K L, WENDEL J F. Polyploidy and genome evolution in plants[J]. *Curr Opin Plant Biol*, 2005b, 8(2):135–141.
- [31] PATERSON A H, CHAPMAN B A, KISSINGER J C, et al. Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon* [J].

- Trends Genet, 2006,22(11):597–602.
- [32] HUGHES T, EKMAN D, ARDAWATIA H, et al. Evaluating dosage compensation as a cause of duplicate gene retention in *Paramecium tetraurelia* [J]. Genome Biol, 2007,8(5):213.
- [33] GOUT J F, DURET L, KAHN D. Differential retention of metabolic genes following whole-genome duplication [J]. Mol Biol Evol, 2009,26(5):1067–1072.
- [34] BYUN-MCKAY S A, GEETA R. Protein subcellular relocalization: A new perspective on the origin of novel genes [J]. Trends Ecol Evol, 2007,22(7):338–344.
- [35] RUIZ F, KRZYWICKA A, KLOTZ C, et al. The *SM19* gene, required for duplication of basal bodies in *Paramecium*, encodes a novel tubulin, η -tubulin [J]. Curr Biol, 2000,10(22):1451–1454.
- [36] RASTOGI S, LIBERLES D A. Subfunctionalization of duplicated genes as a transition state to neofunctionalization [J]. BMC Evol Biol, 2005,5:28.
- [37] PEACOCK C S, SEEGER K, HARRIS D, et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease [J]. Nat Genet, 2007,39(7):839–847.
- [38] CUI J, SMITH T F, SAMUELSON J. Gene expansion in *Trichomonas vaginalis*: A case study on transmembrane cyclases [J]. Genome Inform, 2007,18:35–43.
- [39] NEI M, ROONEY A P. Concerted and birth-and-death evolution of multigene families [J]. Annu Rev Genet, 2005,39:121–152.
- [40] ROONEY A P, WARD T J. Birth-and-death evolution of the internalin multigene family in *Listeria* [J]. Gene, 2008,427(1/2):124–128.
- [41] KEELING P J, DOOLITTLE W F. Concerted evolution in protists: Recent homogenization of a polyubiquitin gene in *Trichomonas vaginalis* [J]. J Mol Evol, 1995,41:556–562.
- [42] PIONTKIVSKA H, ROONEY A P, NEI M. Purifying selection and birth-and-death evolution in the histone H4 gene family [J]. Mol Biol Evol, 2002,19:689–697.
- [43] GARZON-OSPINA D, CADAVID L F, PATARROYO M A. Differential expansion of the merozoite surface protein (*m*sp)-7 gene family in *Plasmodium* species under a birth-and-death model of evolution [J]. Mol Phylogenet Evol, 2010,55(2):399–408.
- [44] KIM S, BACHVAROFF T R, HANDY S M, et al. Dynamics of actin evolution in dinoflagellates [J]. Mol Biol Evol, 2011,28(4):1469–1480.
- [45] NIKOLAIDIS N, NEI M. Concerted and nonconcerted evolution of the Hsp70 gene superfamily in two sibling species of nematodes [J]. Mol Biol Evol, 2004,21(3):498–505.
- [46] NEI M. Gene duplication and nucleotide substitution in evolution [J]. Nature, 1969,221:40–22.
- [47] EISEN J A, COYNE R S, WU M, et al. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote [J]. PLoS Biol, 2006,4(9):e286.
- [48] COYNE R S, HANNICK L, SHANMUGAM D, et al. Comparative genomics of the pathogenic ciliate *Ichthyophthirius multifiliis*, its free-living relatives and a host species provide insights into adoption of a parasitic lifestyle and prospects for disease control [J]. Genome Biol, 2011,12(10):R100.
- [49] DOAK T G, CAVALCANTI A R O, STOVER N A, et al. Sequencing the *Oxytricha trifallax* macronuclear genome: A pilot project [J]. Trends Genet, 2003,19(11):603–607.

Study on the Protists Gene Duplication

HUANG Lijuan, YI Zhenzhen, LIN Xiaofeng*

(Key Laboratory of Ecology and Environment Science in Guangdong Higher Education, College of Life Science, South China Normal University, Guangzhou 510631, China)

Abstract: In this review, we described the studies of two types of gene duplication (large and small-scale gene duplication) and three alternative ways of divergence (neofunctionalization, subfunctionalization and pseudogenization) with protists in detail. Moreover, we introduced the studies of retention bias of duplicated gene and three evolutionary models within protists briefly. It is suggested that more researches should focus on non-model protists in future for promoting the study of protists gene duplication.

Key words: gene duplication; protist; gene preservation; evolutionary models

【中文责编:成文 英文责编:李海航】