

Why are some genomes really big and others quite compact?

浅析基因组大小的进化机制

石米娟, 程莹寅, 张婉婷, 夏晓勤*

中国科学院水生生物研究所, 武汉 430072

* 联系人, E-mail: xqxia@ihb.ac.cn

2016-06-23 收稿, 2016-08-16 修回, 2016-08-18 接受, 2016-08-31 网络版发表

摘要 基因组记录着物种的全套核苷酸序列信息, 其大小与物种进化地位之间的关系由于“C值悖论”的提出而引人关注. 本文回顾了关于基因组大小研究的历史, 简述了基因组大小和细胞性状及自身组成之间关系的研究进展及相关假设, 并在不同分类尺度上分析了影响基因组大小进化的可能因素. 我们认为随着进化地位的提高和相应的遗传信息量增加, 基因组大小有增加的内在必然性; 但是在真核生物中, 随着基因组的增大, 增加的DNA逐渐由基因序列转变为主要是内含子和基因间区等非编码DNA序列, 而非编码序列的大量增加并不明显改变物种的进化地位, 却在很大程度上掩盖了基因组大小与物种进化复杂度之间的相关性, 从而导致了C值悖论. 我们认为分子突变导致了基因组大小的变化, 而自然选择对保留有益的变化起到了重要的作用.

关键词 基因组, 基因组大小, C值悖论, 进化, 基因, 基因间区, 非编码DNA序列

核糖核酸(RNA)和脱氧核糖核酸(DNA)这两类核酸都可以作为遗传信息的载体. 记录生物体自身完整信息的单倍体核酸序列被称之为该物种的基因组, 包括所有的编码及非编码的核酸分子. 广义的基因组包括核基因组、线粒体基因组和叶绿体基因组; 狭义的基因组一般指核基因组. 除RNA病毒外, 其他生物的基因组均由DNA构成. 虽然表观遗传学认为细胞的某些状态(主要是基因组上碱基的各类修饰)也能产生遗传效应, 但确凿无疑的是, 基因组的核酸序列编码了最主要的遗传信息. 虽然二倍体和多倍体生物具有两套或多套在某种意义上等价的遗传信息, 但一套遗传信息即构成该物种的基因组. 基因组的大小称为C值, 该值对一种生物来说是一个常数. 基因组的大小与物种进化地位之间的关系引人关注, 本文在梳理这半个多世纪来人们对基因组大小进化的研究成果的同时, 也介绍了我们在这个问题上的观点.

1 关于基因组大小的科学问题

1.1 基因组大小研究的历史

从20世纪60年代末期开始, 科学家开始对基因组的大小产生了兴趣, 起初的研究一直集中于具有较小基因组的物种, 如病毒和支原体等^[1~3], 后来逐渐扩展至其他多种微生物^[4]、动物^[5,6]和植物^[7], 并对基因组大小变化的机制进行了初步的探索^[8,9], 目前已分别建立了关于动物^[10]和植物^[11]基因组大小的数据库. 在核基因组之外, 叶绿体基因组^[12]和线粒体基因组^[13~16]大小的进化也获得了相应的研究成果. 在这个过程中, 基因组大小的检测技术也在相关的研究中得到发展, 从开始的复性动力学估算法^[17], 到脉冲凝胶电泳法^[18], 到流式细胞法^[19], 再到现在的高通量测序K-mer估计法^[20], 方法越来越简便, 结果越来越准确, 推进着研究的发展.

引用格式: 石米娟, 程莹寅, 张婉婷, 等. 浅析基因组大小的进化机制. 科学通报, 2016, 61: 3188–3195

Shi M J, Cheng Y Y, Zhang W T, et al. The evolutionary mechanism of genome size (in Chinese). Chin Sci Bull, 2016, 61: 3188–3195, doi: 10.1360/N972016-00728

1.2 基因组大小进化的总趋势

生物在进化中, 由分子到细胞, 由原核到真核, 再由单细胞到多细胞, 进而出现组织的分化, 形成器官, 及至由单个个体的生命活动到产生复杂的社群行为. 总体上是由简单到复杂, 由低等到高等, 生命活动形式越来越复杂, 性状也越来越丰富. 生物的性状主要是由遗传信息(主要是基因而非基因间区)决定的, 因此物种的进化必然伴随着越来越多遗传信息的产生. 基因组作为遗传信息的载体, 其长度也不可避免地逐步增加. 事实确实如此, 已有的数据显示基因组的大小同物种的进化程度之间存在一定的正相关关系. NCBI基因组数据库已收录了16092个物种的基因组, 我们按病毒、原核生物和真核生物3大类分别统计不同级别基因组长度的物种数目. 虽然各类的极值多有重叠, 但就均值和中值而言, 病毒<原核生物<真核生物, 基因组大小峰值分属3个不同的区域(图1), 这表明若按这种大尺度分类, 基因组大小和物种复杂程度在总的趋势上是呈正相关性的. 当然, 物种的进化程度不是决定基因组大小的唯一因素, 它仅仅是粗略地反映基因组进化的大趋势, 即复杂的生物需要更大的基因组来承载更为复杂的遗传信息.

1.3 C值悖论

既然基因组记录了一个物种的遗传信息, 人们很自然地把基因组的大小与物种在进化上的复杂程

度相关联起来. 然而随着研究的深入, 人们发现基因组的大小和物种的进化复杂度之间没有严格的对应关系, 这就是所谓的“C值悖论”^[21], 一些物种的基因组大小差异远远大于其进化程度的差异. 比如, 河豚(*Tetraodon nigroviridis*)的基因组大约为340 Mb^[22,23], 而同为硬骨鱼类的非洲肺鱼(*Protopterus aethiopicus*), 其基因组长度为130 Gb, 大380倍以上, 也是人类的基因组(3.2 Gb)的40倍; 从螺旋狸藻(*Genlisea tuberosa*)61 Mb到重楼百合(*Paris japonica*)的150 Gb, 不同被子植物的基因组大小甚至可能相差2400多倍^[24,25]. 令人意外的是, 基因组大小的冠军是单细胞的无恒变形虫(*Polychaos dubium*), 其基因组大小竟然达到惊人的670 Gb^[26]. 此外, 还有一些复杂性相似、甚至表型不可区分的物种, 其基因组大小都可能有极为显著的不同. 如此种种的“异常”, 使人们对“基因组越大, 生物进化越高等”的观点产生了疑问.

2 与基因组大小相关联的因素

2.1 基因组大小与细胞的性状

基因组大小与某些细胞性状存在相关联的现象, 一个极具普遍性的发现是物种的基因组大小和细胞的大小呈正比: $C=kX^{\alpha}$, 其中C为C值, X为细胞大小, k和 α 为常数. 人们发现, 在同一个种属内部, 2个常数时常是固定的^[27]. 随着研究的增多, 基因组大小和细胞大小呈正相关这一论点在很多物种中都得到了证实^[28-30]. 但也有例外, 2008年发表的同科的15种壁

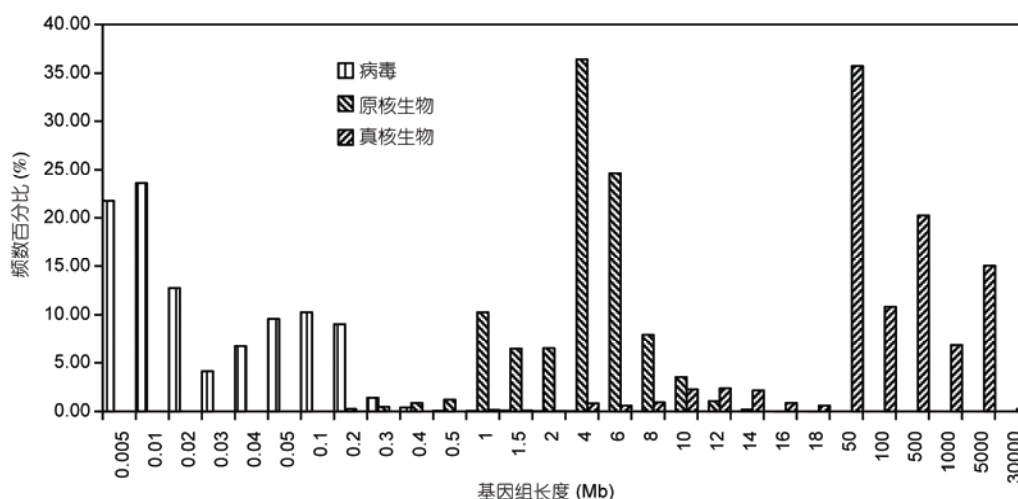


图1 病毒、原核生物和真核生物基因组大小的频数分布百分比

Figure 1 The frequency distributions of genome sizes of viruses, prokaryotes and eukaryotes

虎的血红细胞大小和基因组大小就不存在相关性^[31]. 基于基因组大小和细胞大小的相关性, 研究者们还发现了基因组大小和细胞分裂速率^[32]、胚胎发育速率^[33]、细胞迁移速率等各个性状相关联.

解释这些相关性目前有两类观点^[34]. 第一个观点基于自然选择理论: 生物体在自然选择的压力下朝着“自身利益最优化”的方向发展, 从而使得细胞大小, 甚至其他性状都产生了适应性, 而基因组大小变化则是在这种适应性下的产物, 即细胞大小进化的压力源自于外界. Gregory和Johnston^[35]一直是这一观点的支持者, 他们建立了动物基因组大小的数据库的同时, 也在很多物种中发现了类似的规律, 比如其在果蝇属的55个种的研究中, 发现在18℃的温控下, 不同种的果蝇的发育时长和基因组大小呈正相关. 基于中性进化理论的观点则认为基因组大小的增加是一个“中性突变”的过程, 受制于群体大小, 随着随机的遗传漂变, 使不同物种呈现不同的基因组大小, 而细胞的大小则是为了适应于基因组大小的改变而形成的被动改变, 即细胞大小进化的压力源自于内部.

2.2 基因组大小与基因组的结构组成

目前学术界对于DNA的命名存在不准确的情况, 例如“非编码DNA”(noncoding DNA)通常泛指所有不编码蛋白质的DNA序列, 甚至在不久以前它还当被当成“垃圾”DNA. 虽然不编码蛋白质, 但近些年的研究表明其中有相当一部分可以转录成具有生物学功能的RNA, 仍属于编码RNA分子的基因, 因此称之为“非编码”并不妥当. 为了讨论基因组的结构成分, 我们在以下的讨论中把组成基因组的DNA序列分成基因和基因间区两类. 基因间区完全由非编码DNA序列(noncoding DNA sequence)组成; 基因包括两类, 即编码RNA的基因(RNA-coding gene, R类基因)和编码蛋白质的基因(protein-coding gene, P类基因), 基因序列不仅包括外显子这种编码DNA序列(coding dna sequence), 也包括了相关的非编码DNA序列, 如内含子、启动子和上下游的增强子等调节区域.

基因组大小的变化包括基因大小和数目的变化以及基因间区数量的变化. Lynch等人^[36]曾对基因组大小和外显子DNA、内含子DNA、基因间区DNA在不同物种中的相关性进行了深入的研究, 并在其《基因组结构的起源》一文中进一步阐述. 他发现当基因

组大小在10 Mb以下时, 内含子和基因间区基本不存在, 而在10~20 Mb时, 内含子比例陡增, 当基因组大小在100 Mb以上时, 内含子和外显子的比例基本相同, 而当基因组大小大于2500 Mb, 则绝大多数为内含子和间区序列, 编码DNA的比重跌至10%以下, 甚至1%. 这一统计数据说明, 随着基因组增大, 增多的少部分为编码区, 更多的为非编码区.

2.3 基因组上非编码区的生物学意义

那么非编码区存在的意义何在? 我们认为当基因达到一定数量(比如说2万个左右)之后, 维持生命活动的大部分代谢通路已经基本形成, 进化复杂度已趋于稳定, 进化在分子水平上更多地体现为基因结构与分子互作网络的优化. 在这种情况下, 大量非编码区域一方面贡献了调控元件, 增强了生物分子网络的健壮性; 另一方面还有利于增加物种后代的遗传多样性, 提高物种的环境适应力. 比如大量的内含子可以增加外显子的重组机率, 有利于筛选出更具适应力的外显子组合和基因表型; 同样地, 大量基因间区也可以增加二倍体和多倍体生物后代的基因组合, 从而在基因数量相近的情况下产生更具遗传多样性的后代种群. 内含子和基因间区的增加虽然不明显改变物种在表观上的进化复杂度, 却增强了物种对环境的适应潜力, 或许这就是能适应恶劣干燥环境的非洲肺鱼具有超大基因组的奥秘.

对于物种而言, 基因是细胞行使生理功能的主要分子基础, 它决定了物种的基本特性; 而非编码序列的存在促进了基因的优化、加速了进化历程, 它对物种的影响是通过基因的作用而间接地实现的. 因此我们认为基因的数目应该与物种的进化复杂度直接相关, 而基因间区等非编码序列与物种的进化复杂度之间并没有直接的相关性. 相反, 它们的存在掩盖了基因数目与进化复杂度之间的关系, 使基因组大小看似与进化复杂度无关, 从而产生“C值悖论”.

3 基因组大小变化的分子机制研究

3.1 基因组大小进化与分子突变假说

我们认为不同的因素或机制在不同的尺度上决定了物种基因组大小的进化. 在界一级的层次上, 进化的复杂度决定了基因组的大小发展的方向; 在从门到属以内的层次上, 环境适应性和自然选择起决定作

用;而对于种群间的差异,则可以由中性突变产生。

(i) 随机突变与自然选择在中等尺度上决定了基因组大小的显著变化。在进化历程中基因组上的随机突变可以造成基因组大小的增加或缩减。增加的机制主要是两类: (1) 基因或染色体局部片段的增殖; (2) 整个基因组加倍后异化。而缩减的机制主要是染色体局部片段或整条染色体的丢失。这些事件导致的基因组大小变化是随机的,这些变化在不同环境的选择压力中可能被遗失,但也可能得到巩固。这种基因组大小的变化是随机突变与环境适应的结果,并不与物种复杂度或进化地位相关联,从而可以在物种的门、纲、目、科乃至属的各级分类层级内引起基因组大小的显著不同。

(ii) 中性突变与遗传漂变可以在微尺度上造成基因组大小的波动。中性理论常被认为是继达尔文自然选择理论之后的一个重大的进化理论进展,按照该理论,自然选择对基因组上的大部分突变并不起作用,这些突变通过随机漂移机制引起物种的进化。在我们看来,这个观点是非常片面的——因为不仅从宏观事实上来看,各物种的基因型与表现型无不体现了它们与环境之间的契合,脱离自然选择而仅用遗传漂变理论是无法解释的;而且从逻辑上来讲,对环境选择不敏感、对外在性状都没有明显影响的中性突变不可能对于产生物种特异性性状起多大的作用,难以引起物种的分化,甚至也不对亚种的形成起作用,毕竟亚种也体现了对环境进一步适应。虽然一些分子突变并不能引起明显的表型变化,但变异积累起来就可能引起多基因控制的复杂性状的改变,从而接受自然的选择。正如谢平^[37]所指出的:“所谓中性理论根本不涉及物种分化的问题,而这样的学说怎么可称之为进化理论呢?”

所有的分子突变无非3种作用:有益的突变增加个体对环境的适应能力;有害的突变降低个体的适应能力;而与外在性状无明显关联的中性突变不改变个体的环境适应。事实上在基因组中有大量保守性较低的非编码区域,其中的突变不会造成明显的个体性状差异,难以受到自然选择的作用。因此从分子水平来看,中性理论是有一定依据的,只不过这种中性突变充其量造成物种的不同种群间在基因型上的微小差异,增加物种的遗传多样性,也为有选择意义的表型变化积累分子基础,因此中性突变只能在极微小的尺度上引起基因组大小的变化。应当指出

的是,除中性突变之外,优胜劣汰的自然选择也同样能带来基因组大小的微小变化。

3.2 突变率对基因组大小的影响

对基因组进化的研究,最终需要回到分子生物学的水平求解。普遍观点认为,影响生物基因组大小进化的原因为:突变率(尤其是插入片段与缺失片段的相对频率)和物种群体的大小^[38]。突变率和基因组大小在病毒及原核生物中普遍呈现负相关,而在真核生物中,这一观点存在争议^[39~41]。一般认为,突变为有害突变,若维持高突变率(低保真),则需要繁衍出大规模的群体来降低突变率带来的危害。在这种情况下,如果基因组太大,对生物体的复制而言,是严重的生理消耗。病毒具有高的突变率,使得其基因组一直被限制在一个非常小的范围内,甚至丢弃掉了许多重要的基因,以最小的生理消耗、大规模的种群数量及高的突变率来使其适应复杂的宿主免疫环境。原核生物则具有极保守的突变率,但其由于没有有性生殖带来的缓冲,依然是沿用了病毒应对环境的某些方式,比如小的生理消耗——基因组,及一定规模的种群数量。虽然目前依然不知道具体的碱基突变对基因组大小的影响,但无论是经验还是数据模拟都获得了类似的结论:在无性生殖物种中,突变率和基因组大小呈负相关。

突变带来的危害仅对无性繁殖的物种有着更大的影响,而对有性繁殖物种而言,单碱基的突变很容易被重组等其他生理过程消弭,有性繁殖的物种更容易出现中性突变或者隐性突变,即突变基因至少不会在形成纯合子之前就受到环境的选择。因此对于群体数较小且有性繁殖的大多数真核生物而言,突变率和基因组大小负相关的观点并不适用。事实上,有数据表明真核生物有着比原核生物更高的突变率,这意味着真核基因组的大小有更快的进化速率^[42],并且这种进化速率还会随着基因组的增大而加速^[41]。有趣的是,同一物种的不同细胞突变率也是存在差异的,比如人类的胚胎干细胞、多能干细胞及生殖细胞的突变率就低于体细胞^[43,44]。

3.3 重复序列造成了真核生物基因组大小的巨大差异

绝大多数植物的基因组都经历了多倍化过程,然后经过大规模的染色体重排和删除事件迅速完成

二倍化过程,将基因组的扩大稳定下来^[45,46]。然而真核生物基因组长度的巨大差异主要归因于重复序列(转座因子、卫星DNA与核糖体基因),不仅在种内不同种群间的基因组长度变异主要是由卫星DNA和转座子造成^[47],种间的差异也主要由重复序列引起,其中转座子对于不同种类真核植物的基因组大小的影响尤其显著^[48]。

转座子是造成真核生物基因组大小差异的主要因素,在人类基因组中有近一半的转座子序列,甚至有4%的编码序列中也存在转座子序列^[49],并且这一比例在启动子序列中上升至25%^[50]。若连转座子都可能具有功能,那占人类基因组90%以上不编码蛋白质的“垃圾”DNA又是怎样的情况呢^[51]? ENCODE项目的研究则证明非编码DNA有多方面的作用^[52]。比如后面将提到的piRNA就是非编码DNA的产物,越来越多的研究表明,虽然仅有少数的DNA编码蛋白,但更多的DNA编码着各类的RNA产物,从miRNA, siRNA, 到lncRNA, 再到现在引起关注的circRNA, 从目前公布的人类基因组注释文件来计算,基本有50%以上的序列具有编码RNA的功能。预计这一数字将会随着研究的推进而增长,然而仍然存在大量的非编码区。

3.4 基因组增大的限制机制

基因组也并非是无限制的扩增,巨大的基因组显然会增加生物繁衍复制的成本,如果基因组增大所带来的积极作用不能抵制这种成本,那么进化压力显然会限制基因组大小的进一步增加。虽然并不清楚所有的抑制机制,但有一些已发现的机制是用

于止住基因组扩增步伐的。由RNA干扰相关的沉默通路可以抑制转座子活性来保护基因组^[53,54],这些通路从原核生物到人类中都普遍存在。比如和piwi蛋白家族结合发挥调控作用的piRNA,就可抑制人类精细胞中的转座子活性^[55]。

4 小结与展望

基于对前人的研究工作成果的思考,我们对基因组大小进化问题的认识总结如下:

(1) 生物进化的复杂度增加伴随着基因数量增加和基因组的增大,这是病毒和原核生物、真核生物的基因组大小有大致分区的原因。

(2) 随着基因组的增大,其增大方式由基因数量的增加转变为主要是内含子和基因间区等非编码DNA序列的增加,而非编码序列并不像基因那样明显地改变物种的表观进化复杂度,从而在很大程度上掩盖了基因组大小与进化水平之间的正相关性。

(3) 以转座机制为代表的突变方式是真核基因组的迅速增大的分子基础;分子突变改变基因组的大小,自然选择则决定此种改变是否能长期稳定地遗传下去。

对于分子突变所引起的基因组大小改变是否能稳定遗传下去,我们强调了自然选择在这个过程中的重要作用。基于中性理论的影响力,相信有人仍然会纠结于究竟是自然选择的压力还是遗传漂变决定了基因组大小。无论如何,随着大量物种基因组序列的解析,研究者们已经有越来越多的分子证据来对此进行研究,也许在不久的将来这个问题就可以有广为接受的结论了。

参考文献

- 1 Bak A L, Black F T, Christiansen C, et al. Genome size of mycoplasmal DNA. *Nature*, 1969, 224: 1209–1210
- 2 Askaa G, Christiansen C, Erno H. Bovine mycoplasmas: Genome size and base composition of DNA. *J Gen Microbiol*, 1973, 75: 283–286
- 3 Kelly D C, Avery R J. The DNA content of four small iridescent viruses: Genome size, redundancy, and homology determined by renaturation kinetics. *Virology*, 1974, 57: 425–435
- 4 Pariza M W, Iandolo J J. Determination of genome size of selected typing bacteriophages of *Staphylococcus aureus*. *Appl Microbiol*, 1974, 28: 510–512
- 5 Bachmann K. Genome size in mammals. *Chromosoma*, 1972, 37: 85–93
- 6 Bachmann, K. Genome size and animal evolution. *Riv Istochim Norm Patol*, 1975, 19: 135
- 7 Flavell R B, Bennett M D, Smith J B, et al. Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem Genet*, 1974, 12: 257–269
- 8 Maclean N. Suggested mechanism for increase in size of the genome. *Nat New Biol*, 1973, 246: 205–206
- 9 Sparrow A H, Nauman A F. Evolution of genome size by DNA doublings. *Science*, 1976, 192: 524–529

-
- 10 Gregory T R. Animal genome size database, 2013. <http://www.genomesize.com>
 - 11 Bennett M D, Leitch I J. Plant DNA C-values database, 2012. <http://data.kew.org/cvalues/>
 - 12 Hedberg M F, Huang Y S, Hommersand M H. Size of the chloroplast genome in *codium fragile*. *Science*, 1981, 213: 445–447
 - 13 Khairallah M M, Adams M W, Sears B B. Mitochondrial genome size variation and restriction fragment length polymorphisms in three *Phaseolus* species. *Theor Appl Genet*, 1991, 82: 321–328
 - 14 Boussau B, Brown J M, Fujita M K. Nonadaptive evolution of mitochondrial genome size. *Evolution*, 2011, 65: 2706–2711
 - 15 Rand D M. Endotherms, ectotherms, and mitochondrial genome-size variation. *J Mol Evol*, 1993, 37: 281–295
 - 16 Lagisz M, Poulin R, Nakagawa S. You are where you live: parasitic nematode mitochondrial genome size is associated with the thermal environment generated by hosts. *J Evol Biol*, 2013, 26: 683–690
 - 17 Frenkel N, Roizman B. Herpes simplex virus: genome size and redundancy studied by renaturation kinetics. *J Virol*, 1971, 8: 591–593
 - 18 Chen H, Keseler I M, Shimkets L J. Genome size of *Myxococcus xanthus* determined by pulsed-field gel electrophoresis. *J Bacteriol*, 1990, 172: 4206–4213
 - 19 De Vita R, Cavallo D, Eleuteri P, et al. Flow cytometric approach to study genome size variation in eurasiatic green toads of the *Bufo viridis* complex. *Eur J Histochem*, 1997, 41(Suppl 2): 175–176
 - 20 Chen W, Hasegawa D K, Arumuganathan K, et al. Estimation of the whitefly *Bemisia tabaci* genome size based on k-mer and flow cytometric analyses. *Insects*, 2015, 6: 704–715
 - 21 Callan H G. The organization of genetic units in chromosomes. *J Cell Sci*, 1967, 2: 1–7
 - 22 Roest Crollius H, Jaillon O, Dasilva C, et al. Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome Res*, 2000, 10: 939–949
 - 23 Jaillon O, Aury J M, Brunet F, et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 2004, 431: 946–957
 - 24 Fleischmann A, Michael T P, Rivadavia F, et al. Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Ann Bot*, 2014, 114: 1651–1663
 - 25 Pellicer J, Fay M F, Leitch I J. The largest eukaryotic genome of them all? *Bot J Linn Soc*, 2010, 164: 10–15
 - 26 Parfrey L W, Lahr D J, Katz L A. The dynamic nature of eukaryotic genomes. *Mol Biol Evol*, 2008, 25: 787–794
 - 27 Cavalier-Smith T. Skeletal DNA and the evolution of genome size. *Annu Rev Biophys Bioeng*, 1982, 11: 273–302
 - 28 Gregory T R. The bigger the C-value, the larger the cell: Genome size and red blood cell size in vertebrates. *Blood Cells Mol Dis*, 2001, 27: 830–843
 - 29 Gregory T R. A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class aves. *Evolution*, 2002, 56: 121–130
 - 30 Roth G, Walkowiak W. The influence of genome and cell size on brain morphology in Amphibians. *Cold Spring Harb Perspect Biol* 7, 2015, a019075, doi: 10.1101/cshperspect.a019075
 - 31 Starostova Z, Kratochvil L, Flajshans M. Cell size does not always correspond to genome size: Phylogenetic analysis in geckos questions optimal DNA theories of genome size evolution. *Zoology (Jena)*, 2008, 111: 377–384
 - 32 Chipman A D, Khaner O, Haas A, et al. The evolution of genome size: What can be learned from anuran development? *J Exp Zool*, 2001, 291: 365–374
 - 33 Jalal M, Andersen T, Hessen D O. Temperature and developmental responses of body and cell size in *Drosophila*; effects of polyploidy and genome configuration. *J Therm Biol*, 2015, 51: 1–14
 - 34 Mueller R L. Genome biology and the evolution of cell-size diversity. *Cold Spring Harb Perspect Biol* 7, 2015, doi: 10.1101/cshperspect.a019125
 - 35 Gregory T R, Johnston J S. Genome size diversity in the family Drosophilidae. *Heredity (Edinb)*, 2008, 101: 228–238
 - 36 Lynch M, Koskella B, Schaack S. Mutation pressure and the evolution of organelle genomic architecture. *Science*, 2006, 311: 1727–1730
 - 37 Xie P. Critical Reviews and Reconstruction of Evolutionary Theories (in Chinese). Beijing: Science Press, 2016 [谢平. 进化理论之审读与重塑. 北京: 科学出版社, 2016]
 - 38 Gupta A, LaBar T, Miyagi M, et al. Evolution of genome size in asexual digital organisms. *Sci Rep*, 2016, 6: 25786
 - 39 Sniegowski P D, Gerrish P J, Johnson T, et al. The evolution of mutation rates: Separating causes from consequences. *BioEssays*, 2000, 22: 1057–1066
 - 40 Lynch M. Evolution of the mutation rate. *Trends Genet*, 2010, 26: 345–352
 - 41 Oliver M J, Petrov D, Ackerly D, et al. The mode and tempo of genome size evolution in eukaryotes. *Genome Res*, 2007, 17: 594–601
 - 42 Petrov D A, Sangster T A, Johnston J S, et al. Evidence for DNA loss as a determinant of genome size. *Science*, 2000, 287: 1060–1062
 - 43 TeSlaa T, Setoguchi K, Teitell M A. Mitochondria in human pluripotent stem cell apoptosis. *Semin Cell Dev Biol*, 2016, 52: 76–83

- 44 Hong Y, Cervantes R B, Tichy E, et al. Protecting genomic integrity in somatic cells and embryonic stem cells. *Mutat Res-Fund Mol M*, 2007, 614: 48–55
- 45 Blanc G, Wolfe K H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, 2004, 16: 1667–1678
- 46 Wang X Y, Shi X L, Hao B L, et al. Duplication and DNA segmental loss in the rice genome: Implications for diploidization. *New Phytol*, 2005, 165: 937–946
- 47 Biemont C. Genome size evolution: Within-species variation in genome size. *Heredity (Edinb)*, 2008, 101: 297–298
- 48 Chen J J, Wang Y. Recent progress in plant genome size evolution (in Chinese). *Hereditas (Beijing)*, 2009, 31: 464–470 [陈建军, 王瑛. 植物基因组大小进化的研究进展. *遗传*, 2009, 31: 464–470]
- 49 Nekrutenko A, Li W H. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet*, 2001, 17: 619–621
- 50 Jordan I K, Rogozin I B, Glazko G V, et al. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet*, 2003, 19: 68–72
- 51 Ohno S. So much “junk” DNA in our genome. *Brookhaven Symp Biol*, 1972, 23: 366–370
- 52 ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA elements) Project. *Science*, 2004, 306: 636–640
- 53 Dumesic P A, Madhani H D. Recognizing the enemy within: Licensing RNA-guided genome defense. *Trends Biochem Sci*, 2014, 39: 25–34
- 54 Shabalina S A, Koonin E V. Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol*, 2008, 23: 578–587
- 55 Siomi M C, Sato K, Pezic D, et al. PIWI-interacting small RNAs: The vanguard of genome defence. *Nat Rev Mol Cell Biol*, 2011, 12: 246–258



夏晓勤

1992年毕业于武汉大学生物系; 1997年在中国科学院水生生物研究所获博士学位; 2000年起先后在剑桥大学和美国圣迭戈从事系统生物学、癌症基因组学和生物信息学研究与平台开发; 2011年入选中国科学院“百人计划”研究员, 在水生生物研究所开展水生生物基因组学研究. 研究兴趣涉及细胞代谢网络仿真、高通量组学数据分析、基因芯片数据库与分析平台、鱼类基因组功能注释、草鱼生长性状相关的分子模块耦合分析等方面.

The evolutionary mechanism of genome size

SHI MiJuan, CHENG YingYin, ZHANG WanTing & XIA XiaoQin

Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

A genome is a set of DNA sequences encoding the complete genetic information of an organism, therefore more complex organisms are reasonably expected to have larger genomes. However, eukaryotic genome size vary drastically and seems to have no connection to the evolutionary complexity. Such an inconsistency between them is known as the C-value paradox. Here we briefly introduced the history of researches on genome size, and reviewed the advances and hypotheses of the researches on the relationships between genome size and evolutionary level, genome composition and some cell traits. At the top level of biological classification, the increase of evolutionary complexity is notably coupled with the augment of genetic information which leads to bigger genome sizes. Albeit there are overlapping between adjacent groups, the genome sizes rank upwards as evolutionary complexity, i.e, viruses<prokaryotes<eukaryotes. The genome size correlates with some cell traits, such as cell size, mobility, the rate of cell division, or even the rate of embryo development. One presumption is that the coordinated variation of cell traits and genome size was resulted by evolutionary pressure, and the alternative explanation is that the endogenous growth of DNA led to consequent variations in cell size or other cell traits. The genome composition changes accordingly with the genome size. With the enlargement of eukaryotic genomes, the increments gradually shift from coding DNA sequences to noncoding DNA sequences, which includes introns and intergenic regions. Since genes are the most crucial molecular bases for the physiological functions and gene number determines the complexity of molecular network in a cell, we conjecture that gene number positively correlates to organism's evolutionary complexity. On the other hand, the introduction of noncoding DNA regions provides more *cis*-regulatory element sites, more alternative splicing genotypes, and higher allele exchange frequency, therefore endows a population higher genetic diversity and environmental adaptability. Thus noncoding DNA sequences can be accumulated on genome during evolution because they may promote the optimization of genes and the evolution of an organism. The bulk insertion of noncoding DNA does not apparently change an organism's evolution complexity, but interrupts the correlation between genome size and evolutionary complexity, thereafter leads to the so-called C-value paradox. The evolutionary mechanism of genome size is often explained by two primary hypotheses: the natural selection theory and the neutral theory with genetic drift. We believe the former determines the significant changes in genome size on a mesoscale and the latter contributes to fluctuations in the genome size on a microscale. Molecular events may impose on genome size in diversified ways. While mutation rate and repeated sequences as transposons intend to enlarge genomes, some mechanism, as an example, piRNA can prevent a genome from growing. We emphasize the point that molecular variations bring changes to the genome size and advantageous changes be retained by natural selection.

genome, genome size, C-value paradox, evolution, gene, intergenic region, noncoding DNA sequence

doi: 10.1360/N972016-00728