

DOI: 10.3724/SP.J.1005.2009.00464

# 植物基因组大小进化的研究进展

陈建军, 王瑛

中国科学院武汉植物园, 武汉 430074

**摘要:** 不同的真核生物之间基因组大小差异很大, 并与生物体复杂性不相关, 在基因组中存在大量的非编码 DNA 序列是造成这种差异的主要原因, 特别是转座子序列。文章综述了植物基因组大小差异以及引起这种差异的主要进化动力的最新研究进展。植物基因组多倍化和转座子积累是导致基因组增大的主要动力, 而同源不平重组和非正规重组则是驱动基因组 DNA 丢失的潜在动力, 以制约基因组无限制地增大。文中还讨论了植物基因组大小进化方向, 即总体趋势是朝着增大的方向进化, 某些删除机制主要是削弱这种增大作用但不能逆转。

**关键词:** 基因组大小; 进化动力; 转座子; 非编码 DNA

## Recent progress in plant genome size evolution

CHEN Jian-Jun, WANG Ying

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China

**Abstract:** It has been known that eukaryotic genomes span a wide range of sizes regardless of organism complexity. The observed differences in genome size are primarily due to polyploidy level and abundance of non-coding DNA, especially the contribution of transposable elements (TEs). Here we reviewed the current progress in genome size variation of plant species and the underlying evolutionary forces that contribute to genome expansion or contraction. Polyploidization and the accumulation of transposable element are the primary contributors to genome expansion. As to the mechanisms of DNA loss, unequal homologous recombination and illegitimate recombination are thought to be the counterbalances to the unlimited expansion of a genome. The evolutionary direction of plant genome size is also discussed, which tends to favor larger genomes with deletion mechanisms acting to only attenuate genome expansion but not reverse.

**Keywords:** genome size; evolutionary forces; transposable elements; non-coding DNA

除叶绿体、线粒体等细胞质以外, 植物细胞核是携带绝大部分遗传物质的主要载体。估计细胞核内的 DNA 含量常用 C 值(C-value), 也就是基因组大小来描述<sup>[1]</sup>。C 值是指物种单倍体或配子所含的 DNA 量<sup>[2]</sup>, 单位为 pg(picogram) 或 Mb(million base pair), 而 1 pg 相当于 978 Mb<sup>[3]</sup>。迄今为止, 已建立起含有 5 150 种植物的 DNA C 值数据库(The Plant DNA

C-values Database)<sup>[4]</sup>, 包括被子植物、裸子植物、羊齿类、苔藓类以及藻类等。在被子植物中, 已知最小的基因组发现于螺旋狸藻(*Genlisea maragretae* Hutch.), 基因组大小为 63 Mb<sup>[5]</sup>; 最大的基因组为百合科的四倍体贝母(*Fritillaria assyriaca* Baker.), 基因组大小为 127 Gb<sup>[6]</sup>, 两者相差约 2 000 倍。

研究表明非编码 DNA(Non-coding DNA)是导致

收稿日期: 2008-11-21; 修回日期: 2008-12-23

基金项目: 中国科学院农业基地知识创新工程重要方向项目(编号: KSCX2-YW-N-030), 中科院百人计划项目, 国家自然科学基金项目(编号: 30800624)和武汉市晨光计划项目(编号: 20055003059-45)资助。

作者简介: 陈建军(1979-), 男, 博士, 助理研究员, 研究方向: 植物学。Tel: 027-87510771; E-mail: jianjunchen@wbgcas.cn

通讯作者: 王瑛(1973-), 女, 博士, 研究员, 研究方向: 植物分子遗传学和基因组学。Tel: 027-87510675; E-mail: yingwang@wbgcas.cn

物种间基因组大小差异如此巨大的主要原因<sup>[7]</sup>,然而产生如此众多非编码DNA的进化动力是什么?属于适应性进化吗?如果是,那么非编码DNA在基因组中具有什么样的功能?如果没有适应性进化的作用,那么寄主基因组在自然选择过程中为什么要承载这么多的“垃圾”DNA(Junk DNA)?本文就这些有关基因组大小进化的机制及进化方向等问题进行综述。

## 1 基因组大小的一般性概述

### 1.1 从“C值矛盾”(C-value paradox)到“C值谜”(C-value enigma)

基因组大小与生物体复杂性之间没有关联,即“C值矛盾”<sup>[8]</sup>。起初认为生物体越复杂,进化程度越高,其基因组应该越大;然而,所得到的研究结果却并不一致,例如两栖类动物的基因组比人类的基因组要大。虽然造成“C值矛盾”的主要原因是存在于基因组中非编码DNA所致,但DNA含量变异的起源、作用及意义仍然是个谜。2003年在英国皇家植物园丘园举行的“第二届植物基因组大小研讨会”上,通过了用“C-value enigma”替代“C-value paradox”的新的表述,主要是基于:(1)在真核生物基因组中,非编码DNA有哪些类型,各占的比例有多大?(2)这些非编码DNA来自哪里,并随着时间的推移非编码DNA是怎样在基因组中扩张或丢失的,从而导致物种基因组进化?(3)非编码DNA对染色体、细胞核、细胞及生物体本身有何作用和影响?(4)不同的生物体中,为什么非编码DNA有多有少?据此,以“C值谜”取代“C值矛盾”来表述基因组大小的复杂性更加准确和适用<sup>[9-12]</sup>。

### 1.2 C值的意义和应用

研究物种的基因组大小(C值)为以下几个方面提供理论依据:(1)作为AFLP引物中选择性碱基个数确定的依据,若基因组较大,则选择性碱基相应地增加<sup>[13, 14]</sup>;(2)与微卫星的扩增效率相关,主要是由于大的基因组稀释了靶位点在基因组中所占的比例,致使引物与靶位点结合的几率降低<sup>[15]</sup>;(3)作为细胞核大小、花粉粒直径等表型的预示值,基因组大小与细胞核大小、花粉直径等表型性状成正相关<sup>[16]</sup>;(4)作为生态与环境的预示值,C值与植物的生活史、

地理分布、物候期、单位面积生物量、对生长环境的敏感度(如温度和湿度的变化)等紧密相关,还能预测长期的变化(诸如全球变暖)所引起的植被改变<sup>[17]</sup>;(5)利用推断的C值来揭示古生物学的趋势,例如,根据化石细胞的大小,可以推断化石的DNA含量。在植物化石中,气孔保卫细胞的大小也可以用来估计DNA含量,进而寻求C值的进化趋势<sup>[18]</sup>;(6)C值对物种保育具有重要的意义,大基因组的物种往往存在更大的灭绝风险<sup>[19-21]</sup>。虽然物种灭绝与基因组大小的关系是否存在还不明确,但这种关系还是可能存在的。

### 1.3 C值测定的主要方法

用于测定C值的方法主要有流式细胞分析(Flow cytometry)和孚耳根微显影(Feulgen microdensitometry)。尽管使用孚耳根微显影技术的趋势在减少,但是流式细胞仪技术还不可能完全代替孚耳根微显影技术,在某些方面孚耳根微显影技术也许是更好的选择。使用流式细胞仪时,细胞核是看不到的,染色体数目必须另行细胞学分析,而用孚耳根染色方法可以在同一张玻片上,既可以用于估计C值,又可以检查染色体数目及其倍性<sup>[22]</sup>。目前,两种技术都与计算机成像分析技术系统相结合。

### 1.4 基因组大小与植物表型之间的关系

最初的研究发现基因组大小与细胞大小及减数分裂周期之间存在着正相关<sup>[23]</sup>,此后陆续有研究者提出基因组大小与其他不同层次表型之间的正负相关关系,如基因组大小与细胞大小<sup>[16]</sup>、种子大小之间<sup>[24]</sup>存在着正相关,而与单位面积叶重及代谢率<sup>[25]</sup>等之间存在着负相关。然而,根据Knight等<sup>[26]</sup>最新的统计调查表明,基因组大小与在细胞水平上的表型(如细胞大小,保卫细胞的长度,表皮细胞的面积等)存在着较强的相关性,而与更高尺度上的表型(如单位面积叶重,光合效率等)相关性逐渐减弱。至于基因组大小与表型之间的相关性具有什么样的进化意义目前还不明确。

## 2 基因组大小进化的机制

### 2.1 基因组大小“适应性进化”与“非适应性进化”

从Thomas<sup>[8]</sup>在1971年提出“C值矛盾”以来,围

绕着基因组大小进化过程和基因组复杂性的起源等问题, 众多学者从多个角度提出了许多假设, 也存在很多争议, 争议的焦点主要在于基因组大小是物种“非适应性进化”的结果, 还是“适应性进化”的结果。

Lynch等<sup>[7]</sup>从有效群体大小( $N_e$ )角度来阐述基因组复杂性的起源, 认为从单细胞的原核生物到多细胞的真核生物, 伴随着 $N_e u$ ( $u$ 为突变率)逐渐地减少, 基因组大小却在逐渐地增大。例如, 在辐射状鳍鱼类中, 淡水鱼的有效群体比海鱼要小, 但其基因组却比海鱼的要大, 并且基因组大小与遗传变异、个体大小、世代周期等呈负相关<sup>[27, 28]</sup>。同时有效群体减少很容易固定有害的突变, 一些复制基因、转座子和内含子等得以保留在基因组中, 自然选择的有效性降低, 这是非适应性进化的过程。

然而, 对于Lynch等提出的上述基因组大小“非适应性进化”的假设, 也有人提出质疑, 甚至是截然相反的观点<sup>[29, 30]</sup>, 主要是基于非编码DNA具有功能这一前提, 而不是所谓的“垃圾DNA”。最近Vinogradov等<sup>[31-35]</sup>提出了“genome design”模式: 在温血动物中, 高度表达的基因主要位于富含GC (GC-rich)的区域, 这些区域大都集中在靠近染色质中央, 凝缩程度不高, 易于表达。相反的是, 对于组织特异表达的基因(低表达)多位于缺乏GC (GC-poor)的区域, 这些区域大都在染色质两端, 凝缩程度高, 属于表观调控范围。比较研究发现组织特异表达基因的内含子要比管家基因的内含子长, 这有利于表观遗传调控, 在基因不表达的时候, 由染色质凝缩状态来抑制。同时非编码DNA还可能具有稳定核骨架的作用——调控核质比和细胞大小; 除此以外, 还具有缓冲作用——调节细胞内的能量波动。因此组织特异表达基因的内含子较长与其功能复杂性有关联, 据此, 相当一部分非编码DNA并不是完全中立的, 是适应性进化的结果。因此, 额外的非编码DNA得以保留在物种基因组中, 导致物种基因组大小的进化。

## 2.2 基因组大小增加的进化动力

目前关于基因组大小进化, 多数观点认为与基因组扩张(或DNA含量增加)和删减丢失这两种反方向进化动力的共同作用有关<sup>[36-39]</sup>。对基因组增大扩

张的机制相对比较容易理解, 第一种导致基因组迅速增加的是基因组加倍, 或称为多倍化, 事实上绝大多数植物的基因组都经历过基因组加倍<sup>[40]</sup>, 基因组加倍后再通过大规模的染色体重排或删除事件, 丢失大量复制基因, 迅速完成二倍体化过程。例如在水稻中, 在基因组加倍后有 30%~65%的复制基因丢失<sup>[41]</sup>。尽管植物基因组加倍后还要“缩水”丢失部分DNA, 完成二倍体化过程, 但从总体上来说, 基因组多倍体化仍旧是导致基因组迅速扩张的重要机制。

除了基因组加倍, 导致基因组迅速增加的另一动力是转座子的扩张, 特别是LTR反转座子的扩张, 这在较大的植物基因组中尤为突出<sup>[42]</sup>。在被子植物中, 一般基因组大于 2.5 pg(1C)的物种, 其基因组中重复序列的比例要占到 80%以上, 而这些重复序列又以转座子为主<sup>[43]</sup>。转座子广泛分布于真核生物基因组中, 这些重复序列曾经一度归为“垃圾DNA”, 后来人们发现它们却是产生遗传变异的重要来源, 驱动基因组进化的动力<sup>[44-46]</sup>。目前的研究充分表明, 转座子在影响寄主基因组大小, 介导基因的进化和新基因的形成, 参与基因组的表观调控以及异染色质结构的形成等方面影响着寄主基因组进化<sup>[46-50]</sup>。Rouzic等<sup>[51]</sup>在研究转座子的长期进化时, 发现新的转座子入侵寄主基因组或重新被激活后, 会立即进行有效的转座活动(否则, 会由于自然选择或遗传漂变的作用而被寄主基因组去除), 在短时期内导致寄主基因组的增加, 转座子与寄主基因组发生互作(从致死或有害的转座子插入到寄主基因组逐渐适应这种转座插入), 当寄主基因组接纳新转座子的能力达到某个水平值, 这些转座子也逐渐失去转座活性。依据寄主群体大小, 少数转座子继续保留转座活性, 另一部分转座子被寄主基因组“驯化”, 其余的多数转座子丢失。

在所有植物转座子中, 反转座子是通过“复制-粘贴”的方式来进行转座的, 在寄主基因组中会不断增加自身的拷贝数, 可以在短时期内使得植物基因组大小迅速增加。例如反转座子占玉米基因组的 50%以上<sup>[52]</sup>, 而在过去的 3 百万年之内, 玉米基因组从 1 200 Mb增加到 2 400 Mb, 这主要是由反转座子插入引起的<sup>[53]</sup>。比较玉米和高粱两个亲缘关系较近的物种, 高粱的基因组大小(1.6~1.8 pg/2C)约为玉米基因组大小(4.9~5.5 pg/2C)的三分之一, 尽管两者的基

因组之间存在共线性关系,而分析来自玉米的一段240 kb长的*adh1*侧翼区域时,该区域绝大多数是由LTR反转座子构成,但在高粱基因组中却找不到任何与这些反转座子相关的同源序列。为此,高粱基因组和玉米基因组大小之间的差异是由两个物种从远古共同祖先分化后,LTR反转座子在玉米迅速扩张所导致的<sup>[42, 52, 53]</sup>。另外这种扩张是不均一的,主要是发生在基因间隔区,原因是这些基因间隔区所受到的选择压通常要小<sup>[54]</sup>。

在大基因组植物中许多LTR反转座子家族的拷贝数都超过10 000,而在基因组较小植物中几乎没有超过1 000个拷贝的,说明不同反转座子的扩增潜力是造成被子植物基因组差异的主要因素<sup>[55]</sup>。根据植物中各自的反转座子类型、结构、转座时间和拷贝数等参数,表明大基因组的扩增主要与少数几个反转座子家族有关,不同的物种间具有不同的反转座子家族,而且大部分的反转座子家族都是在近期插入的<sup>[55]</sup>。在褐毛野豌豆(*Vicia pannonica* Crantz.)中,一类*gypsy*反转座子占基因组大小的38%,这类反转座子序列之间极度相似,说明反转座是在近期发生的<sup>[56]</sup>。另外Hawkins等<sup>[57]</sup>深入分析了棉花3个种的基因组大小差异,由拷贝数的估计结果表明该属所含转座子占基因组大小的40%~65%,即约一半以上的基因组由转座子组成。进一步分析表明,*copia*-like转座子主要存在于最小的基因组(*Gossypium raimondii* Ulb.)当中,而*gypsy*-like转座子在比较大的基因组中频繁扩增,一类特别的*gypsy*-like转座子*Gorge3*对基因组大小差异的贡献最大。在野生稻(*Oryza australiensis* Domin.)<sup>[58]</sup>和苕苜(*Lotus japonicus* L.)<sup>[59]</sup>的基因组中也发现了类似的情况。

### 2.3 基因组大小减少的进化动力

相对于基因组增加,导致基因组大小减少最直接的动力还不明晰。最初有研究者意识到indel(insertion-deletion)偏爱会导致基因组DNA丢失,如在蛋白质编码区小片段DNA删除的次数超过同样大小的DNA插入次数<sup>[60]</sup>,而在非编码区,这种DNA删除的片段比插入的片段更频繁<sup>[37]</sup>,这种删除偏爱与滑动复制(Replication slippage)和不等交换(Unequal crossing over)有关,并最终导致DNA丢失。基于indel偏爱的机制,Petrov等<sup>[61]</sup>认为基因组中假

基因(Pseudogenes)的清除效率不同,是导致蟋蟀的基因组比果蝇的基因组大11倍的主要原因。在此基础上逐渐形成所谓的“DNA丢失假设(the DNA loss hypothesis)”,即由indel偏爱所导致的DNA丢失是决定物种间基因组大小不尽相同的原因。然而该假设却受到广泛的置疑<sup>[38, 62]</sup>。尽管对基因组大小进化不起决定性作用,但indel偏爱仍可能是导致物种基因组大小减少的机制之一<sup>[38]</sup>。

另一个基因组DNA删除动力是发生在反转座子之间的重组机制。反转座子短时间内在寄主基因组扩张后,并不会无限地扩张,重组机制可以限制它们对寄主基因组的强烈冲击。寄主基因组会通过不平等同源重组(Unequal homologous recombination)或非正规重组(Illegitimate recombination)而丢失一部分基因组DNA,这些重组主要是通过LTR反转座子两端的LTR序列之间进行重组交换来完成,这使得其中的一些LTR反转座子形成solo LTR反转座子(只留下一端的LTR,其余序列在重组中丢失)。例如在拟南芥基因组中就发现了大量不完整的solo LTR反转座子序列<sup>[63]</sup>,正是通过这些重组机制导致拟南芥基因组DNA的丢失。在水稻<sup>[64-66]</sup>、玉米<sup>[67]</sup>、小麦<sup>[68]</sup>和大麦<sup>[69, 70]</sup>的基因组中也发现了许多结构类似的solo LTR序列。

### 2.4 基因组大小进化的方向

尽管植物基因组大小的进化涉及到基因组增大和减少两个反方向作用力的相互制约,然而物种基因组大小进化的总体趋势到底是朝着哪个方向呢?目前还没有统一的定论。Bennetzen等<sup>[71]</sup>认为除非有足够的证据表明重复DNA能够从基因组中去除,否则植物的基因组只会是越来越大。随后的研究陆续发现通过同源不平等重组或非正规重组等能够将反转座子等重复DNA序列从植物基因组中去除,正是这种反向平衡力的作用限制植物基因组朝着增大的单方向发展进化。然而,通过非正规重组等机制删减的基因组DNA片段很小,从几个碱基对到几百个碱基对不等<sup>[37, 72]</sup>,不足以有效地抵消基因组增加的部分<sup>[73]</sup>。

Leitch等<sup>[6]</sup>从整个陆生植物尺度来分析基因组大小进化,得到基本的趋势是基因组在增大,表明越古老的植物谱系的基因组越小,并逐渐分化出基因组较大的物种。最近Hawkins等<sup>[39, 74]</sup>人认为,从系

统进化的角度来看,基因组大小进化是朝着增大的方向,非正规重组等DNA删除的机制只是起到削弱基因组增大的作用,而不能逆转植物基因组朝着增大方向进化的趋势。例如在稻属中,*O. glaberrima* Steud.是来自非洲的栽培稻,系统进化遗传分析表明,*O. glaberrima*是与*O. sativa*进化关系接近。比较研究发现,与*O. glaberrima*从共同的祖先分离后,由于LTR反转座子的近期爆发,*O. sativa*基因组发生比较明显的扩增(在相同的亚种中,*O. indica*扩大约2%,另一个亚种*O. japonica*基因组扩大约6%)<sup>[75]</sup>,同时在*O. sativa*基因组中也发现通过不平等重组或非正规重组导致大量DNA删除的证据,然而这些DNA删除并不能完全抵消由转座子扩张所引起的寄主基因组增加。换句话说,尽管大量明显的DNA删除反作用力存在,*O. sativa*的基因组大小仍旧是净增长。在棉属物种分化过程中,基因组也是朝着逐渐增大的进化过程<sup>[36, 57, 73, 76]</sup>。

然而现存大多数物种的基因组偏小<sup>[77]</sup>,与上述观点存在矛盾。Hawkins等<sup>[39]</sup>认为大基因组物种在自然选择压的作用下受限制,并且基因组小的物种仍在朝着大的方向进化,不过增加比较缓慢。总之,基于目前所了解的基因组大小进化机制,植物基因组由于转座子爆发等迅速增加,而内在的删除机制和外在选择压只起到削弱延缓增长的作用,植物基因组仍朝着“基因组肥胖的方向(one-way ticket to genomic obesity)<sup>[71]</sup>”进化。

### 3 结语与展望

综上所述,植物基因组大小由于基因组复制或转座子的活跃会在短时期内呈现“爆发”式增长,而转座子爆发大多只涉及到少数几类转座子。基因组内通过 indel 偏爱、不平等同源重组和非正规重组等删除机制反方向制约基因组的扩张,但这种限制作用也十分有限,植物基因组仍可能朝着“基因组肥胖”的方向进化。

为此,我们认为未来的研究需要阐明的是,哪些特殊类型的转座子对促进植物基因组扩张起决定性作用,并且在不同的进化阶段,起决定性作用的转座子从一种类型过渡到另一种类型。同样地,需要进一步证实,基因组删除机制对削减物种基因组的有效性及其程度,将有助于阐明基因组增大和减小之间的动态平衡以及基因组大小最终的进化方

向。另外,从系统进化的角度分析植物基因组大小进化,将为分析阐明植物基因组大小进化的机制和动力提供新的视角。

### 参考文献(References):

- [1] Swift H. The constancy of deoxyribose nucleic acid in plant nuclei. *Proc Natl Acad Sci USA*, 1950, 36(11): 643–654.[\[DOI\]](#)
- [2] Bennett MD, Bhandol P, Leitch IJ. Nuclear DNA amounts in angiosperms and their modern uses-807 new estimates. *Ann Bot*, 2000, 86(4): 859–909.[\[DOI\]](#)
- [3] Dolezel J, Bartos J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. *Cytometry*, 2003, 51(2): 127–128.[\[DOI\]](#)
- [4] Bennett MD, Leitch IJ. Plant DNA C-values database (release 3.0, Dec. 2004). <http://www.kew.org/cvalues/homepage.html>, 2004.
- [5] Greilhuber J, Borsch T, Muller K, Worberg A, Porembski S, Barthlott W. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. *Plant Biol (Stuttg)*, 2006, 8(6): 770–777.[\[DOI\]](#)
- [6] Leitch IJ, Soltis DE, Soltis PS, Bennett MD. Evolution of DNA amounts across land plants (Embryophyta). *Ann Bot*, 2005, 95(1): 207–217.[\[DOI\]](#)
- [7] Lynch M, Conery JS. The origins of genome complexity. *Science*, 2003, 302(5649): 1401–1404.[\[DOI\]](#)
- [8] Thomas CA, Jr. The genetic organization of chromosomes. *Annu Rev Genet*, 1971, 5: 237–256.[\[DOI\]](#)
- [9] Gregory TR. The bigger the C-value, the larger the cell: genome size and red blood cell size in vertebrates. *Blood Cells Mol Dis*, 2001, 27(5): 830–843.
- [10] Gregory TR. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol Rev Camb Philos Soc*, 2001, 76(1): 65–101.[\[DOI\]](#)
- [11] Gregory TR. Genome size and developmental complexity. *Genetica*, 2002, 115(1): 131–146.[\[DOI\]](#)
- [12] Gregory TR. A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class aves. *Evolution*, 2002, 56(1): 121–130.[\[DOI\]](#)
- [13] Fay MF, Cowan RS, Leitch IJ. The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. *Ann Bot*, 2005, 95(1): 237–246.
- [14] Han TH, van Eck HJ, De Jeu MJ, Jacobsen E. Optimization of AFLP fingerprinting of organisms with a large-sized genome: a study on *Alstroemeria* spp. *Theor Appl Genet*, 1999, 98(3–4): 465–471.[\[DOI\]](#)
- [15] Garner TW. Genome size and microsatellites: the effect of nuclear size on amplification potential. *Genome*, 2002, 45(1): 212–215.[\[DOI\]](#)

- [16] Jovtchev G, Schubert V, Meister A, Barow M, Schubert I. Nuclear DNA content and nuclear and cell volume are positively correlated in angiosperms. *Cytogenet Genome Res*, 2006, 114(1): 77–82. [\[DOI\]](#)
- [17] Wakamiya I, Ronald JN, Johnston JS, Price HJ. Genome size and environmental factors in the genus *Pinus*. *Am J Bot*, 1993, 80(11): 1235–1241. [\[DOI\]](#)
- [18] Masterson J. Stomatal size in fossil plants: evidence for poly-ploidy in majority of angiosperms. *Science*, 1994, 264(5157): 421–424. [\[DOI\]](#)
- [19] Vinogradov AE. Selfish DNA is maladaptive: evidence from the plant Red List. *Trends Genet*, 2003, 19(11): 609–614. [\[DOI\]](#)
- [20] Knight CA, Molinari NA, Petrov DA. The large genome constraint hypothesis: Evolution, ecology and phenotype. *Ann Bot*, 2005, 95(1): 177–190. [\[DOI\]](#)
- [21] Vinogradov AE. Genome size and extinction risk in vertebrates. *Proc R Soc B: Biol Sci*, 2004, 271(1549): 1701–1705. [\[DOI\]](#)
- [22] Hardie DC, Gregory TR, Hebert PD. From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. *J Histochem Cytochem*, 2002, 50(6): 735–749.
- [23] Bennett MD. Variation in genomic form in plants and its ecological implications. *New Phytol*, 1987, 106(11): 177–200.
- [24] Beaulieu JM, Moles AT, Leitch IJ, Bennett MD, Dickie JB, Knight CA. Correlated evolution of genome size and seed mass. *New Phytol*, 2007, 173(2): 422–437. [\[DOI\]](#)
- [25] Beaulieu JM, Leitch IJ, Knight CA. Genome size evolution in relation to leaf strategy and metabolic rates revisited. *Ann Bot*, 2007, 99(3): 495–505. [\[DOI\]](#)
- [26] Knight CA, Beaulieu JM. Genome size scaling through phenotype space. *Ann Bot*, 2008, 101(6): 759–766. [\[DOI\]](#)
- [27] Yi S, Streelman JT. Genome size is negatively correlated with effective population size in ray-finned fish. *Trends Genet*, 2005, 21(12): 643–646. [\[DOI\]](#)
- [28] Yi SV. Non-adaptive evolution of genome complexity. *Bio Essays*, 2006, 28(10): 979–982.
- [29] Daubin V, Moran NA. Comment on "The origins of genome complexity". *Science*, 2004, 306(5698): 978. [\[DOI\]](#)
- [30] Vinogradov AE. Testing genome complexity. *Science*, 2004, 304(5669): 389–390. [\[DOI\]](#)
- [31] Vinogradov AE. 'Genome design' model and multicellular complexity: golden middle. *Nucleic Acids Res*, 2006, 34(20): 5906–5914. [\[DOI\]](#)
- [32] Vinogradov AE. "Genome design" model: evidence from conserved intronic sequence in human-mouse comparison. *Genome Res*, 2006, 16(3): 347–354. [\[DOI\]](#)
- [33] Pozzoli U, Menozzi G, Comi GP, Cagliani R, Bresolin N, Sironi M. Intron size in mammals: complexity comes to terms with economy. *Trends Genet*, 2007, 23(1): 20–24. [\[DOI\]](#)
- [34] Sironi M, Menozzi G, Comi GP, Cereda M, Cagliani R, Bresolin N, Pozzoli U. Gene function and expression level influence the insertion/fixation dynamics of distinct transposon families in mammalian introns. *Genome Biology*, 2006, 7(12): R120. [\[DOI\]](#)
- [35] Ren XY, Vorst O, Fiers MW, Stiekema WJ, Nap JP. In plants, highly expressed genes are the least compact. *Trends Genet*, 2006, 22(10): 528–532. [\[DOI\]](#)
- [36] Wendel JF, Cronn RC, Johnston JS, Price HJ. Feast and famine in plant genomes. *Genetica*, 2002, 115(1): 37–47. [\[DOI\]](#)
- [37] Petrov DA. Mutational equilibrium model of genome size evolution. *Theor Popul Biol*, 2002, 61(4): 531–544. [\[DOI\]](#)
- [38] Gregory TR. Insertion-deletion biases and the evolution of genome size. *Gene*, 2004, 324: 15–34. [\[DOI\]](#)
- [39] Hawkins JS, Grover CE, Wendel JF. Repeated big bangs and the expanding universe: Directionality in plant genome size evolution. *Plant Sci*, 2008, 174(6): 557–562. [\[DOI\]](#)
- [40] Blanc G, Wolfe KH. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, 2004, 16(7): 1667–1678. [\[DOI\]](#)
- [41] Wang X, Shi X, Hao B, Ge S, Luo J. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytol*, 2005, 165(3): 937–946. [\[DOI\]](#)
- [42] SanMiguel P, Bennetzen JL. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann Bot*, 1998, 82(Suppl. A): 37–44. [\[DOI\]](#)
- [43] Flavell RB, Bennett MD, Smith JB, Smith DB. Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem Genet*, 1974, 12(4): 257–269. [\[DOI\]](#)
- [44] Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci USA*, 1997, 94(15): 7704–7711. [\[DOI\]](#)
- [45] Biemont C, Vieira C. Junk DNA as an evolutionary force. *Nature*, 2006, 443(7111): 521–524. [\[DOI\]](#)
- [46] Kazazian HH, Jr. Mobile elements: drivers of genome evolution. *Science*, 2004, 303(5664): 1626–1632. [\[DOI\]](#)
- [47] Bennetzen JL. Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol*, 2000, 42(1): 251–269. [\[DOI\]](#)
- [48] Vitte C, Panaud O. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet Genome Res*, 2005, 110(1–4): 91–107. [\[DOI\]](#)
- [49] Brookfield JFY. The ecology of the genome-mobile DNA elements and their hosts. *Nat Rev Genet*, 2005, 6(2): 128–136. [\[DOI\]](#)
- [50] Kidwell MG. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, 2002, 115(1):

- 49–63.[\[DOI\]](#)
- [51] Le Rouzic A, Boutin TS, Capy P. Long-term evolution of transposable elements. *Proc Natl Acad Sci USA*, 2007, 104(49): 19375–19380.[\[DOI\]](#)
- [52] SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL. Nested retrotransposons in the intergenic regions of the maize genome. *Science*, 1996, 274(5288): 765–768.[\[DOI\]](#)
- [53] SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. The paleontology of intergene retrotransposons of maize. *Nat Genet*, 1998, 20(1): 43–45.[\[DOI\]](#)
- [54] Bruggmann R, Bharti AK, Gundlach H, Lai J, Young S, Pontaroli AC, Wei F, Haberer G, Fuks G, Du C, Raymond C, Estep MC, Liu R, Bennetzen JL, Chan AP, Rabinowicz PD, Quackenbush J, Barbazuk WB, Wing RA, Birren B, Nusbaum C, Rounsley S, Mayer KF, Messing J. Uneven chromosome contraction and expansion in the maize genome. *Genome Res*, 2006, 16(10): 1241–1251.[\[DOI\]](#)
- [55] Vitte C, Bennetzen JL. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc Natl Acad Sci USA*, 2006, 103(47): 17638–17643.[\[DOI\]](#)
- [56] Neumann P, Koblikova A, Navratilova A, Macas J. Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. *Genetics*, 2006, 173(2): 1047–1056.[\[DOI\]](#)
- [57] Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res*, 2006, 16(10): 1252–1261.[\[DOI\]](#)
- [58] Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res*, 2006, 16(10): 1262–1269.[\[DOI\]](#)
- [59] Holligan D, Zhang X, Jiang N, Pritham EJ, Wessler SR. The transposable element landscape of the model legume *Lotus japonicus*. *Genetics*, 2006, 174(4): 2215–2228.[\[DOI\]](#)
- [60] de Jong WW, Ryden L. Causes of more frequent deletions than insertions in mutations and protein evolution. *Nature*, 1981, 290(5802): 157–159.[\[DOI\]](#)
- [61] Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL. Evidence for DNA loss as a determinant of genome size. *Science*, 2000, 287(5455): 1060–1062.[\[DOI\]](#)
- [62] Pie M. Evolution of genome size: A phylogenetic test of the DNA loss hypothesis. *Russ J Genet*, 2007, 43(3): 338–340.[\[DOI\]](#)
- [63] Devos KM, Brown JK, Bennetzen JL. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res*, 2002, 12(7): 1075–1079.[\[DOI\]](#)
- [64] Ma J, Devos KM, Bennetzen JL. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res*, 2004, 14(5): 860–869.[\[DOI\]](#)
- [65] Vitte C, Panaud O, Quesneville H. LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss. *BMC Genomics*, 2007, 8: 218.[\[DOI\]](#)
- [66] Vitte C, Panaud O. Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol Biol Evol*, 2003, 20(4): 528–540.[\[DOI\]](#)
- [67] Meyers BC, Tingey SV, Morgante M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res*, 2001, 11(10): 1660–1676.[\[DOI\]](#)
- [68] Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J*, 2001, 26(3): 307–316.[\[DOI\]](#)
- [69] Vicient CM, Suoniemi A, Ananthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell*, 1999, 11(9): 1769–1784.[\[DOI\]](#)
- [70] Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res*, 2000, 10(7): 908–915.[\[DOI\]](#)
- [71] Bennetzen JL, Kellogg EA. Do plants have a one-way ticket to genomic obesity? *Plant Cell*, 1997, 9(9): 1509–1514.
- [72] Bennetzen JL, Ma J, Devos KM. Mechanisms of recent genome size variation in flowering plants. *Ann Bot*, 2005, 95(1): 127–132.[\[DOI\]](#)
- [73] Grover CE, Yu Y, Wing RA, Paterson AH, Wendel JF. A phylogenetic analysis of indel dynamics in the cotton genus. *Mol Biol Evol*, 2008, 25(7): 1415–1428.[\[DOI\]](#)
- [74] Hawkins JS, Hu G, Rapp RA, Grafenberg JL, Wendel JF. Phylogenetic determination of the pace of transposable element proliferation in plants: *copia* and *LINE*-like elements in *Gossypium*. *Genome*, 2008, 51(1): 11–18.[\[DOI\]](#)
- [75] Ma J, Bennetzen JL. Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA*, 2004, 101(34): 12404–12410.[\[DOI\]](#)
- [76] Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF. Incongruent patterns of local and global genome size evolution in cotton. *Genome Res*, 2004, 14(8): 1474–1482.[\[DOI\]](#)
- [77] Oliver MJ, Petrov D, Ackerly D, Falkowski P, Schofield OM. The mode and tempo of genome size evolution in eukaryotes. *Genome Res*, 2007, 17(5): 594–601.[\[DOI\]](#)