

昆虫基因组大小及其进化*

丛宇阳^{1**} 肖花美^{1,2} 李 飞^{1***}

(1. 浙江大学昆虫科学研究所, 杭州 310058; 2. 宜春学院生命科学与资源环境学院,
江西省作物生长发育调控重点实验室, 宜春 336000)

摘 要 基因组是物种遗传信息的集合,其大小是研究基因组进化、结构及功能的重要参数之一。本文介绍了测定基因组大小的方法,简述了基因组大小的进化假说及分子机制,综述了近年来昆虫基因组大小的研究进展,尤其是昆虫基因组大小变化的相关影响因子。总体而言,昆虫基因组大小变化是一个复杂的过程,与转座子的活性有密切的关系,是基因组序列丢失和获得两种过程平衡的结果。基因组大小的变化仍在不断地进化中,其对生物造成的影响是剧烈的,因此对昆虫的表型特征产生了重要的影响,但影响的程度和关系在不同的类群有明显的差异,表现出一定的随机性,目前尚未总结出明显的规律。昆虫是生物多样性最为丰富的动物类群,是研究基因组大小进化的最佳材料,随着越来越多的昆虫基因组被测序公开,对昆虫基因组数据进行深入分析,有利于破解基因组大小进化的“C 值之谜”,可为生物基因组大小的研究提供重要参考。

关键词 昆虫; 基因组大小; C 值悖论; 基因组的进化; 非编码 DNA

Progress in research on insect genome size and evolution

CONG Yu-Yang^{1**} XIAO Hua-Mei^{1,2} LI Fei^{1***}

(1. Institute of Insect Sciences, Zhejiang University, Hangzhou 310058, China; 2. College of Life Sciences and Resources Environment/
Key Laboratory of Crop Growth and Development Regulation, Jiangxi Province, Yichun University, Yichun 336000, China)

Abstract The genome contains the accumulated genetic information of a species, and genome size is an important metric for studying genome structure, function and evolution. Here, we briefly introduce the methods for estimating genome size and the molecular mechanisms and evolutionary forces underlying changes in animal genome size. We also review current progress in studying insect genome size, especially the correlation between genome size and variable phenotypic characteristics in insects. In general, evolutionary change in insect genome size is a quite complex process that is closely related to transposon. Genome size is the result of the balance between DNA loss and gain. Insect genome size is still evolving, which in turn influences the phenotypic characteristics of insects to various degrees, however, the explicit relationship between these two features remains obscure. Since they have the highest diversity of all animal taxa, insects are good models for studying genome size evolution. Further analysis of insect genome data facilitates uncovering evolutionary changes in insect genome size, which may also shed light on the “C-value enigma”.

Key words insect; genome size; C-value paradox; genome evolution; non-coding DNA

基因组大小 (Genome size, GS) 是指成熟细胞内单倍体 DNA 含量的总和,常用 C 值 (C-value) 表示,单位为 pg (Picogram), 1 pg 的基因组约等于 978 Mb (Million base pair) 的

DNA 序列。目前,动物基因组大小数据库 (Animal genome size database) 提供了超过 8 000 种动物的基因组大小的数值,其中包括 1 300 条的昆虫基因组大小记录,涵盖捻翅目、纺足目、石蛎目

*资助项目 Supported projects: 国家自然科学基金 (31772238; 31760514)

**第一作者 First author, E-mail: conguyang@zju.edu.cn

***通讯作者 Corresponding author, E-mail: lifei18@zju.edu.cn

收稿日期 Received: 2019-08-20; 接受日期 Accepted: 2019-09-20

等 21 个目。迄今为止, 基因组最小的昆虫为双翅目的海滨摇蚊 *Clunio tsushimensis*, 基因组大小为 0.07 pg (约 68.5 Mb), 基因组最大的为直翅目的斑腿蝗 *Podisma pedestris*, 基因组大小为 16.93 pg (约 16.6 Gb), 两者相差 242 倍。

昆虫基因组具高杂合、高重复的特点, 重复序列是导致不同昆虫基因组产生巨大差异的主要原因之一。近年来, 随着基因组数据不断积累, 许多研究者从不同角度对昆虫基因组大小与一些表型特征进行了关联分析, 同时也给基因组大小进化的不同假说提供了证据。为此, 本文对基因组大小的测定、昆虫基因组大小的影响因素、进化动力以及昆虫基因组大小进化的分子机制等多个领域进行了综述。

1 昆虫基因组大小与影响因素分析

1.1 基因组大小的测定及分析方法

评估基因组大小 (Genome size estimation, GSE) 是研究特定物种基因组的重要前提, 是基因组测序和组装的重要参考指标。目前常用的评估昆虫基因组大小的方法主要有两种, 一种是通过流式细胞仪 (Flow cytometry) 分析得到细胞染色体组型的信息并推断其基因组大小, 另一种是进行低深度基因组测序, 然后采用 K-mer 法 (He *et al.*, 2016) 来预测基因组大小、杂合度及重复序列等信息。近年来, 为提高数据分析质量和效率, 开发了对两种方法获得的数据进行基因组大小分析的方法和软件, 如基于流式细胞仪输出数据进行相关分析的 flowPloidy (R 包) (Smith *et al.*, 2018), 基于 K-mer 频次数据进行分析的 findGSE (R 包) (Sun *et al.*, 2018), GenomeScope (Vurture *et al.*, 2017), gce 等。利用流式细胞仪的方法需要基因组大小已知的物种作为标准对照, 如预测昆虫基因组大小时, 通常采用黑腹果蝇 *Drosophila melanogaster* 作为参考。研究发现, 不同实验室环境下饲养的黑腹果蝇, 其基因组大小存在高达 0.05 pg (约 48.9 Mb) 的差异 (Palmer *et al.*, 2003)。因此, 除实验误差影响外, 评估基因组大小也会受“标

准参照”的差异而产生较大偏差。

1.2 昆虫基因组大小的影响因素

早在 1971 年, Thomas 等 (1971) 即提出了著名的“C 值悖论 (C-value paradox)”: 真核生物的基因组大小变化范围巨大, 而基因组大小的变化与生命体自身的复杂性并无关联。昆虫种类繁多, 不同的昆虫类群有着变化多样的表型特性和生活史特征, 但很多特性与昆虫基因组大小并无明显的关联。例如, 对捻翅目 (Johnston *et al.*, 2004), 膜翅目 (Ardila-Garcia *et al.*, 2010) 和蜚蠊目 (Koshikawa *et al.*, 2008) 中一些具有寄生性或社会性的昆虫种类的研究发现, 其基因组大小在系统发育上与这些昆虫固有的这些特性并不相关; 与寡食性昆虫相比, 杂食性昆虫对复杂植物种类的次生代谢产物有着更强的分解能力 (Baroncelli *et al.*, 2016), 但通过对几种食性明确已知的近缘棉铃虫的基因组大小与其食性在系统发育进化上的相关性分析, 发现昆虫的食性与基因组大小并无显著的关系 (Zhang *et al.*, 2019)。

另一方面, 由于基因组大小与细胞核大小和细胞大小等细胞水平的特征紧密相关, 生物细胞数量和大小变化可能导致体型大小的变化, 昆虫基因组大小与昆虫的体型大小存在潜在的相关性。将基因组大小与体型大小进行关联分析, 发现体虱 (Johnston *et al.*, 2007), 蚜虫 (Finston *et al.*, 1995), 伊蚊 (Ferrari and Rai, 1989) 和蚂蚁 (Tsutsui *et al.*, 2008) 的基因组大小与体型大小的相关性并不显著, 而拟步甲属 (Palmer and Petitpierre, 1996; Palmer *et al.*, 2003) 几种昆虫的体型大小与其基因组大小呈较强的负相关性。环境温度的变化对生物体体型大小有显著影响 (The temperature-size rule), 尤其是在变温动物发育后期, 环境温度的增加将导致其最终体型的减小 (Hessen *et al.*, 2013)。生物的发育和生存环境的变化密切相关, 直接通过基因组大小“预测”体型大小, 不具有合理性。此外, 也有研究发现基因组大小与雄性昆虫的精子发育、受精成功率、雌性昆虫的生殖力呈正相关 (Tsutsui

et al., 2008; Arnqvist *et al.*, 2015), 但与昆虫变态发育特征的相关性并不明显 (Gregory, 2002)。

继 2000 年黑腹果蝇基因组计划的完成, 近年来基因组测序成本大幅度降低, 越来越多的昆虫基因组被揭开神秘的面纱, 研究人员将基因组进行注释分析, 发现不同昆虫种类之间的基因组结构特征相似性与差异性并存。例如, 基因组大小超过 6 Gb 的蝗虫 *Locusta migratoria* (Wang *et al.*, 2014), 其基因组中可编码蛋白的基因总数为 17 000 多个, 而基因组大小不过 100 Mb 的海滨摇蚊, 其编码基因数量却超过 20 000 个 (Kaiser *et al.*, 2016), 这表明编码基因数量与两物种间的基因组大小并不相关。随着一个个物种基因组测序计划的完成, 以转座子为代表的非编码 DNA 在基因组中的重要地位慢慢被挖掘。在生物入侵新地区并长期定殖下来的过程中, 与环境压力相关的转座子转座活性将增强, 对基因组大小的变化易产生潜在影响 (Capy *et al.*, 2000; Kim *et al.*, 2014; Piacentini *et al.*, 2014)。对分布在不同地理区域的黑腹果蝇的研究发现, 长期定殖于发源地以外的地理种群, 其基因组大小显著大于发源地种群, 发源地以外的地理种群之间的基因组大小差异并不显著 (Vieira *et al.*, 2002; Nardon *et al.*, 2005), 这说明昆虫的入侵性和定殖能力与基因组大小也密切相关。

2 基因组大小的进化假说: 选择性假说和遗传漂变

基因组大小在生命体的长期进化过程中受自然选择的压力或遗传漂变过程的影响而被固定下来 (Petrov, 2001)。

基因组大小进化的“选择性假说”强调自然选择在基因组的进化中起主导作用, 该理论认为生物在环境压力的选择下, 其基因组大小会被调试到一个合适的最优值, 在此过程中, 基因组的配置元件不断更新, 细胞水平的表型特征 (如细胞大小、细胞代谢速率等) 发生改变, 生物个体因此从中“受益” (Gregory, 2001)。这一选择压力表现为生存栖境的变化, 例如在平均温度

低、降水量强的地区中生存的几种盘腹蚁属 *Aphaenogaster* 的昆虫倾向于有较大的基因组 (Lau *et al.*, 2019); 也体现在雌雄虫交配时的偏好反应, 例如雄性蝗虫通过“歌唱”的方式求偶, “歌声”具有吸引力的雄虫, 其基因组显著小于“资质平平”的雄虫, 而更受雌虫青睐, 这说明性选择在基因组大小进化中也具有重要的作用 (Schielzeth *et al.*, 2014)。选择压力可以作用于基因组的分子元件, 也可作用于细胞, 共同推动基因组大小的进化。在基因组进化的分子水平上, 转座子的迅速累积导致的基因组“肥胖”对宿主基因组本身既可以产生中性或有害的影响, 也可以产生有益的影响 (Chuong *et al.*, 2017)。在多细胞生物中, 较小的基因组细胞代谢能力强, 例如具有飞行能力的鸟类与其他哺乳动物相比, 其基因组偏小 (Wright *et al.*, 2014), 而鳞翅目昆虫中, 具有较强迁飞能力的帝王蝶为了适应飞行, 在长期的选择压力下, 基因组更为紧缩 (Gregory and Hebert, 2003)。另外, 核内遗传物质含量少的生物其细胞增殖发育的周期循环会更迅速, 较小的代谢速率可以使生物消耗很少的能量就能在生态环境中占据一个合适的生态位 (Vinogradov, 2004)。除此之外, 精细胞也通过减小核基因组的内容量以增加精子的数量, 从而促进生物体内的“生殖竞争”, 筛选最优的精子 (Gage and Morrow, 2003)。

然而, 基因组大小进化的“遗传漂变”假说则指出基因组大小最初在谱系中便被固定下来, 与选择压力无关。基因组复杂性的增加是由于在长期进化的过程中, 生命个体的体型不断增大, 所在种群其有效种群大小 (N_e) 不断减小, 在起主导作用的非适应性随机遗传漂变过程中, 基因组发生变化, 从而引起基因组大小的变化 (Lynch and Conery, 2003)。这一理论强调有效种群大小的重要性, 但对鞘翅目几种豆象的基因组大小与有效种群大小的研究中, 未发现两者存在直接的关联性 (Arnqvist *et al.*, 2015)。

3 基因组大小进化的分子机制

在生物进化的过程中, 一些基因组的组成元

件容易发生活跃的变化。这些变化作用于基因组的局部或整体,对后续基因组的转录调控至关重要。由于基因组元件发生变化而累积的分子突变导致 DNA 的丢失 (DNA loss) 和获得 (DNA gain), 是个体水平基因组大小进化的原动力, 主要包括碱基的插入和缺失变异、转座子转座导致基因组的结构变异、卫星序列的扩张与收缩以及染色体变异等。

3.1 碱基的随机插入和缺失变异 (Insertion and deletion-Indel)

将不同生命个体的基因组水平内的相似序列进行比较,发现某些碱基出现空位或冗余,序列的插入和缺失 (DNA loss and gain) 是解释这一现象的主要成因。仅次于单碱基替换,碱基的随机插入或缺失在基因组的进化过程中的发生频率很高,是形成基因组复杂而多样化的主要动力 (Fan *et al.*, 2007)。不同物种基因组的插入和缺失的速率不同。夏威夷竹节虫 *Laupala molokoensis* 基因组的 DNA 缺失速率比黑腹果蝇低 40 倍,是导致其基因组大小远超果蝇 11 倍的主要原因 (Petrov *et al.*, 2000)。由于碱基的局部或全基因组范围内插入缺失变异对真核生物的进化具有普适性和重要性, Petrov (2002) 提出了基因组大小进化的“突变平衡”模型: 由于随机插入和缺失事件的存在, DNA 缺失和获得速率不同,产生突变偏差,这种偏差在选择压力下被负选择作用所消除或被正选择作用所放大,进而导致亲缘关系较近的物种 (Hjelmen *et al.*, 2017) 或同一物种的雌雄个体 (Hjelmen *et al.*, 2019) 间基因组大小存在较大差异。基于这一假说,有研究指出,在果蝇基因组大小的进化过程中, DNA 缺失事件是导致果蝇基因组大小变化的决定因素 (Assis and Kondrashov, 2012; Ellis *et al.*, 2014)。

3.2 转座子 (Transposable elements, TE) 的插入 (增殖) 或剪切 (抑制) 事件

转座子是一种广泛存在于真核生物体内,可以在基因组中自由跳跃的 DNA 序列。通过“剪切-粘贴”(DNA 转座子)或“复制-粘贴”(反转

录转座子)的方式随机地在基因组的基因区和基因间区跳跃 (Boeke *et al.*, 1985)。这种可移动的基因组元件通过改变基因组的结构对基因进行遗传调控,进而可能对生物体的形态特征和生长发育产生重大影响 (Doolittle, 2013)。在已测序和注释的昆虫基因组中,转座子是含量最丰富的 DNA 组分,尤其体现在基因组较大的物种中。有着 6.3 Gb DNA 含量的东亚飞蝗,其基因组中转座子的含量高达 60% (Wang *et al.*, 2014),德国小蠊 *Blattella germanica* (基因组大小 2 Gb) 和堆沙白蚁 *Cryptotermes secundus* (基因组大小 1 Gb) 中所含转座子也均占基因组大小的一半以上 (Harrison *et al.*, 2018)。

目前,关于转座子与基因组大小进化之间的关联存在不同的假说争议。“自私的基因”假说认为: 在进化过程中,核 DNA 面临最直接的选择压力是可以在细胞中生存并繁衍复制,大量存在的转座子的“自私”表现为: 必定存在一些种类的转座子“无表型相关性 (No phenotypic)”。此类转座子不与任何表型和进化相关的功能相联系,其存在仅是为了繁衍而复制自身的 DNA 序列,以此提高基因在遗传上的多样性 (Genetic variability) 和在细胞中的存活率,进而加速基因组变异和多元化的进程 (Orgel and Crick, 1980)。经典的自然选择假说则认为: 自然选择的压力影响与物种适合度相关的转座子,转座子的活性随着生物对环境的适应而不断变化,这类被环境压力调动起来的分子元件在基因组中或跳跃或粘贴,从而导致在进化过程中物种基因组大小发生改变 (Gregory and Hebert, 1999; Smit, 1999; Kidwell and Lisch, 2001; Biemont and Vieira, 2006)。2000 年以来,随着转座子研究热度的不断升温,研究人员以果蝇属昆虫为研究对象,对转座子与基因组大小的关系进行了大量研究 (Vieira *et al.*, 2002; Boulesteix *et al.*, 2006; Sessegolo *et al.*, 2016), 这些研究从不同角度为基因组大小进化的自然选择假说提供了证据。

3.3 微卫星序列 (Microsatellite) 的扩张和收缩

微卫星序列是一段包含 1-6 个碱基的简单重

复片段 (Simple sequence repeats , SSR) , 广泛存在于真核生物的基因组中 , 其不同重复拷贝类别在种属及个体间存在特异性 , 是基因组序列的重要组成部分 (Ellegren , 2004) 。 微卫星序列广泛分布于染色体的异染色质区 (如端粒和着丝粒) 。 Bosco 等 (2007) 对果蝇染色体异染色质区的微卫星序列的研究表明 , 微卫星序列的扩张和收缩可能是导致其种间和种内的基因组大小呈多样性的主要原因。此外也有研究发现 , 在黑腹果蝇和冈比亚按蚊 *Anopheles gambiae* 这两种 XY 性别决定型的昆虫中 , Y 染色体的大部分区域都是低转录的异染色质区 , 具有大量的重复序列 , 因此由 Y 染色体进化导致的基因组重复序列的扩张与收缩 , 可能与物种基因组大小的进化相关 (Bonaccorsi *et al.* , 1980) , 同种生物雌雄两性基因组的大小之间的差异可能也与此相关。

3.4 染色体的变异

多倍体化是染色体数量变异的一种 , 由于多倍体化而导致的全基因组加倍是基因组大小倍数级差异的主要原因 , 在植物中较为常见 (Blanc and Wolfe , 2004) 。 自然界中 , 多倍体动物较为少见 , 大部分昆虫均为二倍体 , 单倍体或三倍体只在某些特殊物种 (如鞘翅目象甲科 *Curculionidae* 的某些象甲、双翅目毛蠓科的毛蠓 *Psychoda parthenogenetica*) 中出现 (Lachowska *et al.* , 2008) 。 然而 , 昆虫不同组织细胞中存在的染色体核内多倍性 (Endopolyploidy) 却非常常见 , 某些组织可能由于行使特殊功能的需要 , 染色体发生了数量级的加倍。比如家蚕丝腺中的核 DNA 经过 30 次左右的复制形成了“ 100 万倍体 ” (Perdixgillot , 1979) , 这是在长期的选择压力下单个细胞基因组大小特化的体现。此外 , 在某些直翅目蝗虫中也发现存在正常核型范围内的超数染色体 (B chromosome) (Jones and Rees , 1982) , B 染色体中含有大量的重复序列 (Trivers *et al.* , 2004) , 其数量变化与某些物种基因组大小存在关联 (Schielzeth *et al.* , 2014) 。

上述分子突变在进化的选择压力下不断累积 , 导致基因组的结构发生变化 , 主要体现在基

因组整体的 GC 含量 (Bohlin *et al.* , 2014 ; Li and Du , 2014 ; Almpanis *et al.* , 2018) 。 基因编码区某些基因家族的扩增与收缩 (Gardiner *et al.* , 2008 ; Hessen *et al.* , 2013) 、 内含子长度 (Moriyama *et al.* , 1998) 等 , 这些基因组的结构特征也直接或间接地与基因组大小存在相关性。另一方面 , 细胞和染色体作为基因组的载体 , 核 DNA 量的变化也会影响细胞核型、染色体数目、细胞大小以及细胞代谢速率等细胞特征。有研究指出蜻蜓目 (Ardila-Garcia and Gregory , 2009) 、 缨尾目 (Jacobson *et al.* , 2013) 、 石蛎目 (Gassner *et al.* , 2014) 、 鞘翅目 (Petitpierre *et al.* , 1993) 和膜翅目 (Gokhman *et al.* , 2017) 中的某些昆虫 , 其染色体数目或平均长度均与其基因组大小相关 , 而双翅目果蝇 (Gregory and Johnston , 2008) 和半翅目蚜虫 (Finston *et al.* , 1995) 的基因组大小与染色体数目却无明显的相关性。基因组大小与细胞水平的表型存在着较强的相关性。基因组的扩张可能导致细胞转录产物的增加 , 细胞通过调节自身的大小 (比表面积) 以维持代谢过程中氧气和废物运输的平衡 , 细胞在一定范围内变大 , 分裂速度变慢 , 细胞周期可能更长 (Bennett , 1971) 。

4 展望

昆虫作为物种丰富度最为繁盛的动物类群 , 其分布范围广 , 对环境适应性强 , 是基因组大小进化问题的最佳研究对象。综上所述 , 昆虫基因组大小的变化是一个复杂的过程 , 是基因组序列丢失和获得的平衡 , 与转座子和微卫星序列的扩增和收缩有密切的关系。基因组大小的变化是一个剧烈的过程 , 因此对昆虫的表型特征产生了重要的影响 , 但影响的程度和关系在不同的类群有明显的差异 , 目前尚未总结出明显的规律特征 , 表现出一定的随机性。

随着多种昆虫基因组测序的完成 , 越来越多的昆虫基因组数据公开发表 , 我们将可以对各类基因组的组成元件的变化进行更为细致的研究 , 探明昆虫基因组中各组成元件的结构和功能 , 系

系统地阐释这些组成元件的进化影响昆虫基因组大小变化的作用, 有利于破解基因组大小进化的“C 值之谜”。其中, 基因组中非蛋白编码区域的进化包括重复序列、非编码 RNA 等是未来研究的重点。

参考文献 (References)

- Almpanis A, Swain M, Gatherer D, McEwan N, 2018. Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. *Microbial Genomics*, 4(4): 000168.
- Ardila-Garcia AM, Gregory TR, 2009. An exploration of genome size diversity in dragonflies and damselflies (Insecta: Odonata). *Journal of Zoology*, 278(3): 163–173.
- Ardila-Garcia AM, Umphrey GJ, Gregory TR, 2010. An expansion of the genome size dataset for the insect order Hymenoptera, with a first test of parasitism and eusociality as possible constraints. *Insect Molecular Biology*, 19(3): 337–346.
- Arnqvist G, Sayadi A, Immonen E, Hotzy C, Rankin D, Tuda M, Hjelmen CE, Johnston JS, 2015. Genome size correlates with reproductive fitness in seed beetles. *Proceedings of the Royal Society B: Biological Science*, 282(1815): 20151421.
- Assis R, Kondrashov AS, 2012. A strong deletion bias in nonallelic gene conversion. *PLoS Genetics*, 8(2): e1002508.
- Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, Le Floch G, Harrison RJ, Holub E, Sukno SA, Sreenivasaprasad S, Thon MR, 2016. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics*, 17(1): 555.
- Bennett MD, 1971. The duration of meiosis. *Proceedings of Royal Society B: Biological Sciences*, 178(1052): 277–299.
- Biemont C, Vieira C, 2006. Genetics-junk DNA as an evolutionary force. *Nature*, 443(7111): 521–524.
- Blanc G, Wolfe KH, 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, 16(7): 1667–1678.
- Boeke JD, Garfinkel DJ, Styles CA, Fink GR, 1985. Ty elements transpose through an RNA intermediate. *Cell*, 40(3): 491–500.
- Bohlin J, Sekse C, Skjerve E, Brynildsrud O, 2014. Positive correlations between genomic %AT and genome size within strains of bacterial species. *Environmental Microbiology Reports*, 6(3): 278–286.
- Bonaccorsi S, Santini G, Gatti M, Pimpinelli S, Colluzzi M, 1980. Intraspecific polymorphism of sex chromosome heterochromatin in two species of the *Anopheles gambiae* complex. *Chromosoma*, 76(1): 57–64.
- Bosco G, Campbell P, Leiva-Neto JT, Markow TA, 2007. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics*, 177(3): 1277–1290.
- Boulesteix M, Weiss M, Biemont C, 2006. Differences in genome size between closely related species: The *Drosophila melanogaster* species subgroup. *Molecular Biology and Evolution*, 23(1): 162–167.
- Capy P, Gasperi G, Biemont C, Bazin C, 2000. Stress and transposable elements: co-evolution or useful parasites? *Heredity (Edinb)*, 85 (2): 101–106.
- Chuong EB, Elde NC, Feschotte C, 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*, 18(2): 71–86.
- Doolittle WF, 2013. Is junk DNA bunk? A critique of ENCODE. *Proceedings of the National Academy of Sciences of the United States of America*, 110(14): 5294–5300.
- Ellegren H, 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5(6): 435–445.
- Ellis LL, Huang W, Quinn AM, Ahuja A, Alfjred B, Gomez FE, Hjelmen CE, Moore KL, Mackay TFC, Johnston JS, 2014. Intrapopulation genome size variation in *D. melanogaster* reflects life history variation and plasticity. *PLoS Genetics*, 10(7): e1004522.
- Fan Y, Wang W, Ma G, Liang L, Shi Q, Tao S, 2007. Patterns of insertion and deletion in Mammalian genomes. *Current Genomics*, 8(6): 370–378.
- Ferrari JA, Rai KS, 1989. Phenotypic correlates of genome size variation in *Aedes Albopictus*. *Evolution*, 43(4): 895–899.
- Finston TL, Hebert PDN, Footitt RB, 1995. Genome size variation in aphids. *Insect Biochemistry and Molecular Biology*, 25(2): 189–196.
- Gage MJ, Morrow EH, 2003. Experimental evidence for the evolution of numerous, tiny sperm via sperm competition. *Current Biology*, 13(9): 754–757.
- Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG, 2008. *Drosophila* chemoreceptor gene evolution: Selection, specialization and genome size. *Molecular Ecology*, 17(7): 1648–1657.
- Gassner M, Dejaco T, Schonswetter P, Marec F, Arthofer W, Schlick-Steiner BC, Steiner FM, 2014. Extensive variation in chromosome number and genome size in sexual and parthenogenetic species of the jumping-bristletail genus *Machilis*

- (Archaeognatha). *Ecology Evolution*, 4(21): 4093–4105.
- Gokhman VE, Kuhn KL, Woolley JB, Hopper KR, 2017. Variation in genome size and karyotype among closely related aphid parasitoids (Hymenoptera, Aphelinidae). *Comparative Cytogenetics*, 11(1): 97–117.
- Gregory TR, 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews*, 76(1): 65–101.
- Gregory TR, 2002. Genome size and developmental complexity. *Genetica*, 115(1): 131–146.
- Gregory TR, Hebert PDN, 1999. The modulation of DNA content: Proximate causes and ultimate consequences. *Genome Research*, 9(4): 317–324.
- Gregory TR, Hebert PDN, 2003. Genome size variation in Lepidopteran insects. *Canadian Journal of Zoology-revue Canadienne De Zoologie*, 81(8): 1399–1405.
- Gregory TR, Johnston JS, 2008. Genome size diversity in the family Drosophilidae. *Heredity (Edinb)*, 101(3): 228–238.
- Harrison MC, Jongepier E, Robertson HM, Arming N, Bitard-Feildel T, Chao H, Childers CP, Dinh H, Doddapaneni H, Dugan S, Gowin J, Greiner C, Han Y, Hu HF, Hughes DST, Huylmans AK, Kemena C, Kremer LPM, Lee SL, Lopez-Ezquerria A, Mallet L, Monroy-Kuhn JM, Moser A, Murali SC, Muzny DM, Otani S, Piulachs MD, Poelchau M, Qu JX, Schaub F, Wada-Katsumata A, Worley KC, Xie QL, Ylla G, Poulsen M, Gibbs RA, Schal C, Richards S, Belles X, Korb J, Bornberg-Bauer E, 2018. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution*, 2(3): 557–566.
- He K, Lin KJ, Wang GR, Li F, 2016. Genome sizes of nine insect species determined by flow cytometry and k-mer analysis. *Frontiers in Physiology*, 7: 569.
- Hessen DO, Daufresne M, Leinaas HP, 2013. Temperature-size relations from the cellular-genomic perspective. *Biological Reviews*, 88(2): 476–489.
- Hjelmen CE, Garrett MA, Holmes VR, Mynes M, Piron E, Johnston JS, 2019. Genome size evolution within and between the sexes. *Journal of Heredity*, 110(2): 219–228.
- Hjelmen CE, Johnston JS, 2017. The mode and tempo of genome size evolution in the subgenus *Sophophora*. *PLoS ONE*, 12(3): e0173505.
- Jacobson AL, Johnston JS, Rotenberg D, Whitfield AE, Booth W, Vargo EL, Kennedy GG, 2013. Genome size and ploidy of Thysanoptera. *Insect Molecular Biology*, 22(1): 12–17.
- Jones RN, Rees H, 1982. B Chromosomes. Chicago: Academic Press. 1–266.
- Johnston JS, Ross LD, Beani L, Hughes DP, Kathirithamby J, 2004. Tiny genomes and endoreduplication in Strepsiptera. *Insect Molecular Biology*, 13(6): 581–585.
- Johnston JS, Yoon KS, Strycharz JP, Pittendrigh BR, Clark JM, 2007. Body lice and head lice (Anoplura: Pediculidae) have the smallest genomes of any hemimetabolous insect reported to date. *Journal of Medical Entomology*, 44(6): 1009–1012.
- Kaiser TS, Poehn B, Szkiba D, Preussner M, Sedlazeck FJ, Zrim A, Neumann T, Nguyen LT, Betancourt AJ, Hummel T, Vogel H, Dorner S, Heyd F, von Haeseler A, Tessmar-Raible K, 2016. The genomic basis of circadian and circalunar timing adaptations in a midge. *Nature* 540(7631): 69–73.
- Kidwell MG, Lisch DR, 2001. Perspective: Transposable elements, parasitic DNA, and genome evolution. *Evolution*, 55(1): 1–24.
- Kim YB, Oh JH, McIver LJ, Rashkovetsky E, Michalak K, Gamer HR, Kang L, Nevo E, Korol AB, Michalak P, 2014. Divergence of *Drosophila melanogaster* repeatomes in response to a sharp microclimate contrast in Evolution Canyon, Israel. *Proceedings of the National Academy of Sciences of the United States of America*, 111(29): 10630–10635.
- Koshikawa S, Miyazaki S, Cornette R, Matsumoto T, Miura T, 2008. Genome size of termites (Insecta, Dictyoptera, Isoptera) and wood roaches (Insecta, Dictyoptera, Cryptocercidae). *Naturwissenschaften*, 95(9): 859–867.
- Lachowska D, Rozek M, Holecova M, 2008. New data on the cytology of parthenogenetic weevils (Coleoptera, Curculionidae). *Genetica*, 134(2): 235–242.
- Lau MK, Ellison AM, Nguyen A, Penick C, DeMarco B, Gotelli NJ, Sanders NJ, Dunn RR, Cahan SH, 2019. Draft *Aphaenogaster* genomes expand our view of ant genome size variation across climate gradients. *Peer J*, 7: e6447.
- Li XQ, Du D, 2014. Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. *PLoS ONE*, 9(2): e88339.
- Lynch M, Conery JS, 2003. The origins of genome complexity. *Science*, 302(5649): 1401–1404.
- Moriyama EN, Petrov DA, Hartl DL, 1998. Genome size and intron size in *Drosophila*. *Molecular Biology and Evolution*, 15(6): 770–773.
- Nardon C, Weiss M, Vieira C, Biemont C, 2003. Variation of the genome size estimate with environmental conditions in *Drosophila melanogaster*. *Cytometry Part A*, 55A(1): 43–49.
- Orgel LE, Crick FHC, 1980. Selfish DNA-the ultimate parasite. *Nature*, 284(5757): 604–607.
- Palmer M, Petitpierre E, 1996. Relationship of genome size to body size in *Phylan semicostatus* (Coleoptera: Tenebrionidae). *Annals*

- Entomological Society of America*, 89(2): 221–225.
- Palmer M, Petitpierre E, Pons J, 2003. Test of the correlation between body size and DNA content in *Pimelia* (Coleoptera: Tenebrionidae) from the Canary Islands. *European Journal of Entomology*, 100(1): 123–129.
- Perdrixgillot S, 1979. DNA-synthesis and endomitoses in the giant nuclei of the silk gland of *Bombyx mori*. *Biochimie*, 61(2): 171–204.
- Petitpierre E, Segarra C, Juan C, 1993. Genome size and chromosomal evolution in leaf beetles (Coleoptera, Chrysomelidae). *Hereditas*, 119(1): 1–6.
- Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL, 2000. Evidence for DNA loss as a determinant of genome size. *Science*, 287(5455): 1060–1062.
- Petrov DA, 2001. Evolution of genome size: New approaches to an old problem. *Trends in Genetics*, 17(1): 23–28.
- Petrov DA, 2002. Mutational equilibrium model of genome size evolution. *Theoretical Population Biology*, 61(4): 531–544.
- Piacentini L, Fanti L, Specchia V, Bozzetti MP, Berloco M, Palumbo G, Pimpinelli S, 2014. Transposons, environmental changes, and heritable induced phenotypic variability. *Chromosoma*, 123(4): 345–354.
- Schielzeth H, Streitner C, Lampe U, Franzke A, Reinhold K, 2014. Genome size variation affects song attractiveness in grasshoppers: Evidence for sexual selection against large genomes. *Evolution*, 68(12): 3629–3635.
- Sessegolo C, Burlet N, Haudry A, 2016. Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. *Biology Letters*, 12(8): 20160407.
- Smit AFA, 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Current Opinion in Genetics & Development*, 9(6): 657–663.
- Smith TW, Kron P, Martin SL, 2018. FlowPloidy: An R package for genome size and ploidy assessment of flow cytometry data. *Applications in Plant Sciences*, 6(7): e01164.
- Sun H, Ding J, Piednoel M, Schneeberger K, 2018. FindGSE: Estimating genome size variation within human and Arabidopsis using k-mer frequencies. *Bioinformatics*, 34(4): 550–557.
- Thomas CA, 1971. The genetic organization of chromosomes. *Annual Reviews of Genetics*, 5(1): 237–256.
- Trivers R, Burt A, Palestis BG, 2004. B chromosomes and genome size in flowering plants. *Genome*, 47(1): 1–8.
- Tsutsui ND, Suarez AV, Spagna JC, Johnston S, 2008. The evolution of genome size in ants. *BMC Evolutionary Biology*, 8(1): 64.
- Vieira C, Nardon C, Arpin C, Lepetit D, Biemont C, 2002. Evolution of genome size in *Drosophila*. Is the invader's genome being invaded by transposable elements? *Molecular Biology Evolution*, 19(7): 1154–1161.
- Vinogradov AE, 2004. Evolution of genome size: Multilevel selection, mutation bias or dynamical chaos? *Current Opinion Genetics & Development*, 14(6): 620–626.
- Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, 2017. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, 33(14): 2202–2204.
- Wang XH, Fang XD, Yang PC, Jiang XT, Jiang F, Gurtowski J, Schatz MC, 2014. The locust genome provides insight into swarm formation and long-distance flight. *Nature Communications*, 5(1): 1–9.
- Wright NA, Gregory TR, Witt CC, 2014. Metabolic 'engines' of flight drive genome size reduction in birds. *Proceedings of the Royal Society B: Biological Science*, 281(1779): 20132780.
- Zhang S, Gu S, Ni X, Li X, 2019. Genome size reversely correlates with host plant range in *Helicoverpa* species. *Frontiers in Physiology*, 10: 29.