脊椎动物多倍体基因组演化研究现状——模型选取及研究方法进展

柴静^{1*}, 林国亮¹, 黄峰¹, Robert W. Murphy^{2,3}

1. 云南大学生命科学学院, 省部共建云南省生物资源保护与利用国家重点实验室, 昆明 650091;

2. 中国科学院昆明动物研究所, 遗传资源与进化国家重点实验室, 昆明 650223;

3. Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto M5S 2C6, Ontario, Canada

* 联系人, E-mail: jingchai313@foxmail.com

收稿日期: 2019-02-01; 接受日期: 2019-03-08; 网络版发表日期: 2019-04-15

国家自然科学基金(批准号: 91631305, 31760306)、2016年度“博士后创新人才支持计划”(批准号: BX201600130)和第61批中国博士后科学基金面上项目(批准号: 2017M613015)资助

摘要 多倍化即全基因组加倍的现象, 在植物中普遍存在, 而在脊椎动物中相对罕见. 多倍体植物类群如何克服“基因组休克”效应达成基因组稳定及演化等问题均有系统的研究和成熟的理论, 同时也为脊椎动物多倍化研究提供了理论参考. 本文综述了多倍化研究中的理论发展, 以及脊椎动物类群中的研究模型和进展, 论述了组学研究面临的挑战以及有效的技术策略, 以期能为未来解析脊椎动物多倍体演化问题提供切实可行的思路.

关键词 脊椎动物, 多倍化, 组学, 研究模型

多倍化(polyloidization)是指染色体组加倍的现象, 也就是全基因组复制(whole genome duplication)^[1]. 多倍化通过合并和加倍基因组, 从而使生物体的遗传物质结构同时发生加倍和重塑^[2]. 在自然界中, 引发多倍化的可能途径有两种: 通过自身全基因组加倍或近缘物种的基因组融合的方式称为同源多倍化(auto-polyploidization); 通过杂交或远缘基因组融合后发生基因组加倍的方式, 即为异源多倍化(allopolyploidization)^[1,3]. 全基因组复制现象为生物体提供了更多的遗传物质, 同时也会引发基因组冗余^[4].

在自然界中, 多倍化在大部分的植物、真菌、低等动物, 甚至低等脊椎动物中均存在. 现今地球上生存

的约60%~70%开花植物都为多倍体, 或可能有多倍体祖先^[5,6], 特别是一些与人类生产生活息息相关的重要作物, 如小麦、油菜、咖啡、棉花等均为多倍体^[7~9]. 多倍化被推测与将近2%~4%的物种形成分化事件有关^[10]. 而在脊椎动物中, 多倍体现象相对较为罕见, 主要存在于两爬类和硬骨鱼类的部分物种中^[11~13].

1 多倍化现象概述及其研究理论发展

1.1 植物多倍体研究提示初期变化决定了物种能否成功幸存

多倍化发生之后, 冗余的遗传物质首先会引发遗

引用格式: 柴静, 林国亮, 黄峰, 等. 脊椎动物多倍体基因组演化研究现状——模型选取及研究方法进展. 中国科学: 生命科学, 2019, 49: 484–497
Chai J, Lin G L, Huang F, et al. Genomic evolution in polyploid vertebrates: Models and progress in research techniques (in Chinese). Sci Sin Vitae, 2019, 49: 484–497, doi: [10.1360/N052018-00210](https://doi.org/10.1360/N052018-00210)

传水平的变化,无论在同源多倍体还是异源多倍体中,全基因组加倍以及不同来源的基因组相融合都可能会引起全基因组范围的转座元件(transposable elements, TE)脱抑制,进而引起广泛的遗传结构改变^[14].该过程类似于McClintock^[15]所描述的“基因组休克”效应(genome shock)的变化,于是多倍化研究者借用了这个概念来描述物种全基因组加倍之后的变化,包括基因组不稳定性、染色体配对不平衡以及调控不相容性等^[1].基于植物多倍体的研究提示,基因组休克效应在细胞学水平、遗传水平、表达调控以及表观修饰等层面^[16-18].这些变化的综合效应,具体可能包含了基因的随机丢失、突变的加速形成、染色体重排以及同源染色体配对失败等^[19,20].冗余的遗传物质可能导致细胞构架和细胞内调控网络的改变,这可能包括转录水平、转录后水平的变化以及DNA甲基化和核仁显性等,如同源基因位点修饰方式的改变,以及基因的表达规律变化^[21-25].表观遗传水平的改变可能是可逆的,同时也可能导致如基因沉默和活化等变化由最初的可逆进一步转变为不可逆的遗传特征^[16,26].这些现象也可能导致非正常的表达变化和剂量不平衡效应等方面的基因组水平紊乱^[1].另外,在生命体的配子和受精卵等发生过程中,全基因组复制还可能导致不正常的减数分裂或有丝分裂进程.增加的基因剂量可能会引起疾病的发生和不正常的发育,也决定了其能否存活^[10,27].

基于多种植物多倍化体系的研究均提示,初期阶段基因组休克引发的综合效应,可能是多倍体是否能转向稳定而存活的关键.相对于古老的多倍体而言,一些已知遗传背景的新产生的多倍体(如一百年内)成为研究多倍化发生初期基因表达调控等关键改变的理想材料.追踪多倍化发生初期遗传和表观遗传水平的关键改变,为揭示多倍化发生及幸存的可能机制提供线索.在开花植物中,多倍化的相关研究相对动物而言已经比较成熟,基于植物多倍体的遗传和表观遗传水平的研究可为动物多倍体研究的开展提供借鉴.

1.2 多倍化现象在脊椎动物中罕见的原因

多倍化除了引发基因组结构的改变之外,在表型上也会引起许多改变,这些改变诸如植物中植株的育性、杂交活力以及开花时间等^[28].这些表型形状可能使得多倍体植株生物量增长,生长更迅速、更具活力,甚至产生新的性状、发展新功能并占据新的生态

位^[29-31].

相对于多倍化现象在植物中的普遍性,这一现象在动物中较为罕见,目前仅存于昆虫、硬骨鱼类和两栖爬行类的少数物种中^[11-13].为何多倍化现象在脊椎动物中罕见呢?目前已有的假设认为,基因组加倍可能引发了基因组休克效应及剧烈的基因组重塑,从而造成脊椎动物中的性别决定、生理和发育过程中的一系列障碍,以至于多倍体后代在短期世代的胚胎早期便不能存活^[32,33].在异源四倍体鱼类研究中发现,多倍体体系初期世代中产生的大量部分同源基因重组、DNA片段丢失和新基因的形成,可能导致基因组水平剧烈的重塑过程^[34].除以上原因之外,更多用以解释为何多倍化在脊椎动物中较为罕见的原因或证据还有待提出.再者,从遗传水平、表观遗传水平和基因的表达调控网络改变等层次来看,哪一个层次发生的变化为主导,或者哪些变化的相互作用为关键,这些问题均有待于深入研究.

1.3 选取合适的研究模型是阐明多倍化现象罕见原因的关键

植物多倍化体系的大量研究为脊椎动物的研究提供了很好的示例.基于天然多倍化群体的研究发现,基因组休克效应引发的基因组重塑如染色体重排和基因重组等在幸存的多倍体子代中得到了保留;其次,基因型变异与表型、生态分布的关联性分析等,为寻找和筛选产生新形状的突变提供了线索^[8,35].基于人工繁育的多倍化体系研究,如棉花、芸薹属植物等,均提示了初期世代大量的基因组重塑事件,基于拟南芥的研究更是全面地阐明了遗传和表观遗传水平的剧烈变化^[9,19,20,36,37].根据这些研究可知,选取合适的研究模型是研究多倍化各层次问题的关键.在脊椎动物中,在组学技术应用发展之前,由于缺乏合适的研究模型,许多研究仅停留在依据少量数据的间接证据、数学建模模拟推测等^[10,38].选取合适的物种或体系作为研究对象,对多倍化后遗传和表观遗传水平的变化深入研究,是阐明多倍化现象在动物中罕见原因的关键.

1.4 脊椎动物中潜在的研究模型

(1) 脊椎动物中存在的天然多倍体和人工培育体系为潜在的良好模型.在脊椎动物中,多倍化现象主要集中分布于两栖爬行类和硬骨鱼类的少数物种

中^[13,39]。近年来报道的一些脊椎动物也可能经历了多次全基因组复制事件, 在其中一些类群, 多倍化现象可能还在持续发生^[40,41]。

(i) 两栖爬行类中的多倍体分布。在两栖类动物中, 许多属或种内存在多种倍性或基因组大小变化的现象。雌性*Bufo baturae*(=*Bufo pseudoraddei*, $2n=3X=33$)能同时进行有性和无性繁殖, 提示这一物种中存在特殊的调控减数分裂的潜在能力。这些蟾蜍为研究多倍体基因重组和非重组的进化驱动力提供了很好的材料^[42]。除蟾蜍外, 在叶水蛙多倍体的研究中, Gruber和Silva等人^[43]通过分析二倍体(*Phyllomedusa distincta*, $2n=26$)和四倍体(*P. tetraploidea*, $2n=4X=52$)以及二者天然杂交后代($2n=3X=39$)的染色体, 获得四倍体中染色体组型发生二倍化趋势的证据。

非洲爪蟾属的一些物种被用于脊椎动物胚胎发育等研究。在该属中, 存在四倍体(如*Xenopus vastitus*, *X. laevis*, $2n=4X=72$)和六倍体(如*X. ruwenzoriensis*, $2n=6X=108$)物种^[44]。其中四倍体爪蟾*X. laevis*起源于二倍体物种*X. tropicalis*, 与其二倍体祖先物种相比, 该四倍体物种在恶劣环境中具备极强的抗逆性、抗病性, 具有很强的入侵性。该四倍体物种还能产生54条染色体的三倍体后代, 并能通过雌核发育产生36条染色体的二倍体后代。三倍体后代为研究发育过程中细胞核与细胞质间的互作以及性别分化问题提供研究模型^[45]。

在爬行纲动物中, 多个属或种也存在多倍体物种, 如蜥蜴中, *Aspidoscelis*和*Darevskia*属中存在三倍体物种^[46-48], 目前认为这些三倍体可能是由于精子进入加倍的卵细胞中从而形成三倍体个体, 如二倍体物种*A. neomexicana*和二倍体物种*A. sexlineata viridis*杂交形成相应的三倍体, 该三倍体中含有*A. neomexicana*全部染色体和一套来自*A. sexlineata viridis*的染色体^[47]。在*Darevskia*属中的二倍体*D. armeniaca*和*D. unisexualis*会与雌雄同体的二倍体*D. valentine*发生杂交, 从而产生三倍体*D. armeniaca*和*D. unisexualis*^[48]。

(ii) 硬骨鱼类中的多倍体分布。硬骨鱼类是目前脊椎动物中最为繁盛的一个类群, 现存物种数量已超过28000种, 占脊椎动物物种数量的1/2以上, 是脊椎动物中在形态学、生态学、生理学、行为及生活史等方面最具多样性的类群^[10,49]。该类群中有许多物种可能经历了多次全基因组复制事件^[50-52]。在辐鳍鱼纲中,

鲱形目、鲤形目、鲑形目以及胭脂鱼科均报道了多倍体鱼类。在这些类群中, 鲱科(基因组大小为1.22~9.32 pg不等)、鲑科(1.80~4.90 pg)和鲤科(0.80~11 pg不等)鱼类存在多种不同基因组大小的物种(www.genomesize.com), 并且存在种间甚至种内的倍性多样性现象^[39]。例如, 鲑科中的大鳞大麻哈鱼(*Oncorhynchus tshawytscha*)具有二倍体($2n=68$)和三倍体($2n=3X=102$)两种倍性的个体^[53]; 虹鳟鱼(*Salmo gairdneri*)的二倍体染色体数目为59或60, 三倍体染色体数目为89或90^[54]。而鲤形目中, 多倍化现象可能参与了其中约120个属的演化历史, 可能包含了400余个近缘多倍体物种^[51,55,56]。此外, 与被子植物多倍化群体类似, 脊椎动物中一些较为近期的全基因组复制事件可能由杂交引起。例如, 在*Tropidophoxinellus alburnoides*复合种中, 存在二倍体($2n=50$)和三倍体($2n=3X=75$)种, 四倍体则可能是由三倍体种繁育而来^[57]。

(2) 鲤科鱼类可作为研究脊椎动物古多倍化的理想模型。在鲤科鱼类中, 多倍化现象相对集中分布于鲤亚科(Cyprininae)、鲃亚科(Barbinae)、裂腹鱼亚科(Schizothoracinae)和野鲮亚科(Labeoninae)中, 这四个亚科包含了超过120个属的鱼类, 广泛分布于欧亚大陆南部及非洲的淡水水域(www.FishBase.com)^[51,55,56], 具备了基因组大小、染色体数目最为多态的多倍体类群, 并发生了继3R之后的多次同源和异源多倍化。其中约30个属包含了大量的四倍体($2n=4X=100$)和六倍体($2n=6X=150$ ~)物种^[51,55]。更有甚者, 分布于青藏高原腹地及周边地区的裂腹鱼亚科鱼类, 包含了染色体数目为60~424之多的物种^[58]。这几个亚科中最为人们熟知的是鲤亚科鱼类, 其中包括常见的淡水鱼类鲤鱼和鲫鱼, 是目前多倍化起源历史解析得比较清楚的代表类群。基于线粒体基因组和多个核基因的系统发育关系构建发现, 鲫属和鲤属鱼类可能起源于10~12个百万年(Ma)前的异源多倍化事件^[40], 后经历基因组二倍化。其中, 分布于云南的天然鲫鱼(*Carassius auratus*)复合种存在四倍体($2n=4X=100$)、六倍体($2n=6X=150$ ~168)甚至八倍体($2n=8X=200$)三种倍性的群体, 基于线粒体和核基因的系统发育关系构建, 推测该类群鱼类可能经历了同域分布下反复发生的同源多倍化和异源多倍化事件^[59]。并且, 六倍体群体在高海拔和高污染的水域中大量存在, 提示这一复合种鱼类可能是用于阐明多倍体发生及其对极端环境适应等

问题的理想模型^[60,61]. 此外, 鲫属鱼类中的红鲫(*Carassius auratus* var. red)可与不同属(如鲤鱼, *Cyprinus carpio*, $2n=100$)甚至不同亚科(如团头鲂, *Megalobrama amblycephala*, $2n=48$)的鱼杂交, 并获得倍性稳定遗传的异源多倍体后代^[62,63].

综上, 鲤科鱼类表现出的基因组高可塑性和对全基因组复制的非凡耐受力, 具备作为古多倍体基因组演化以及多倍体群体基因组演化历史研究模型的优势.

(3) 杂交多倍化体系的建立为研究多倍化初期变化提供了便利. 利用红鲫和鲤鱼进行远缘杂交获得的鲫鲤杂交体系可能是研究脊椎动物多倍化发生初期基因组各水平快速变化的最佳模型之一. 这一体系是由湖南师范大学刘少军研究团队人工繁育所得. 该团队利用红鲫(♀, $2n=4X=100$)和鲤鱼(♂, $2n=4X=100$)进行属间远缘杂交, 获得的前两代为四倍体(F_1 , F_2 , $2n=100$), 由 F_2 代产生的100条染色体的精子、卵子相结合, 获得八倍体 F_3 代($2n=8X=200$), 经连续自交, 获得了倍性稳定遗传的八倍体世代群体($F_3 \sim F_{26}$, $2n=8X=200$), 这一杂交和多倍化体系的建立仅仅经历了约40年时间^[62,63]. 其次, 为了探索类似鲫鲤杂交的反交——鲤鲫杂交组合中是否也能选育出异源八倍体鱼, 该团队利用锦鲤(♀, $2n=4X=100$)和花鲫(♂, $2n=4X=100$)进行属间远缘杂交获得鲤鲫杂交体系, 杂交 F_1 代的鱼中只存在一种倍性($2n=4X=100$). 通过对鲤鲫 F_1 进行人工催产并使它们自交, 获得了存活的四倍体和六倍体鲤鲫 F_2 代, 这说明鲤鲫 F_1 中存在可育的雌性和雄性个体^[64]. 以上杂交鱼的正、反交体系构建, 为比较杂交和多倍化发生初期, 脊椎动物在遗传变异、表达调控网络改变、基因印记以及甲基化水平改变等表观遗传修饰变化提供了理想模型.

此外, 刘少军团队还利用亚科间远缘杂交技术, 构建了多个远缘杂交体系. 例如, 鲫鲂(红鲫(♀, $2n=100$) $\times$ 团头鲂(♂, $2n=48$))^[63,65]、鲫鲮(红鲫(♀, $2n=100$) $\times$ 黄尾密鲮(♂, $2n=48$))^[66]、鲫鲃(红鲫(♀, $2n=100$) $\times$ 翘嘴红鲃(♂, $2n=48$))^[67]、鲂鲃(团头鲂(♀, $2n=48$) $\times$ 翘嘴红鲃(♂, $2n=48$))^[68,69]、鲂鲮(团头鲂(♀, $2n=48$) $\times$ 黄尾密鲮(♂, $2n=48$))^[70]和草鲂(草鱼(♀, $2n=48$) $\times$ 团头鲂(♂, $2n=48$))^[71]等杂交体系.

2 多倍体组学研究中的技术发展、难点与挑战

2.1 组学的发展为研究遗传和表观遗传水平改变提供便利

在人类基因组公布之时, 已有多倍体植物的基因组陆续报道和公布, 如拟南芥(*Arabidopsis thaliana*, 2000), 水稻中的籼稻(*Oryza sativa* L. ssp. Indica)^[72]和粳稻(*Oryza sativa* L. ssp. japonica)^[73]等. 随着组学测序和分析技术的发展和技术的降低, 无论是古多倍化事件的追踪, 还是新生多倍体在“基因组休克”影响下, 物种在遗传、基因表达和调控以及表观遗传水平的变化, 均得到了全面和深入的阐述.

在基因组结构研究方面, 通过全基因组遗传成分、共线性的比较以及重组事件的挖掘, 可以得知古多倍化事件所引发的结构变异. 例如, 基于高通量测序数据分析获得的棉花基因组中, 通过对异源四倍体棉花A, D两个亚基因组的比较, 发现A亚基因组大规模的转座元件扩张、染色体重排事件, 而D亚基因组则发生了大量的缩减^[74]; 其次, 通过序列变异特征的追踪, 以及演化速率的变化规律, 提示了棉花基因组中的非对称的部分同源基因重组事件^[9,20,75]. 在芸薹属植物中, Liu等人^[19]通过比较两个姐妹物种甘蓝(*Brassica oleracea*)和芜菁(*B. rapa*)的基因组, 发现了包括非对称式基因丢失、特异通路中不同的基因共保留现象, 此外还有转座元件的非对称式扩张以及大规模的染色体重排等, 更为清楚地阐述了多倍体基因组的遗传结构改变.

在基因表达调控变化研究方面, 在古多倍体植物研究中, 通过比较多倍体和二倍体植物的表达量差异, 提示了多倍体植物在表型上优于二倍体祖先物种的原因. 例如, 棉花A, D亚基因组在四倍体中的功能分工以及相较于二倍体基因的表达优势^[74,75]. 芸薹属植物中, 多倍化后同源基因对在表达量上体现功能分化^[18], 且发生亚功能化和新功能化的基因受到更强的选择压力^[76]. 另一方面, 通过对植物人工多倍化体系大量的转录组分析, 提示了多倍化初期世代中由“基因组休克”进一步所引发的“转录组休克”现象和细节^[77]. 例如, 在拟南芥杂交体系中, 多倍体子代同源基因表现出基因的叠加和非叠加式表达^[28], 这些表达量的叠加、激活或抑制, 可能体现了基因组融合和加倍之后

的调控网络剧变, 这其中可能是表观修饰以及转录后调控机制的综合结果^[28,78]。

2.2 脊椎动物中的多倍化研究

(1) 脊椎动物演化早期的多倍化证据. 多倍化在脊椎动物中的研究可追溯至1970年, Ohno^[79]在其论著*Evolution by Gene Duplication*中提出了基因重复(gene duplication)的概念. 他通过研究脊椎动物的DNA含量发现, 有颌脊椎动物是无脊椎动物的四倍, 由此而提出了2R假说, 也就是说, 在由无脊椎动物到有脊椎动物的演化历程中, 可能发生了两次全基因组复制事件. 他认为, 随机的基因复制以及全基因组复制可能是脊椎动物发生辐射演化的重要驱动力之一. 随后, 基于Hox(Homeobox)基因簇的一系列研究发现, 处于无脊椎动物与无颌脊椎动物之间的过渡动物文昌鱼(*Amphioxus*)只有一个Hox基因簇, 而无颌动物代表七鳃鳗(*Lamprocybus*)则拥有两个或者三个; 从鱼类到人类中, 均存在四个或更多个独立的Hox基因簇^[50]. 该研究极大支持了Ohno的观点. 后来, Furlong和Holland^[80]基于同工酶的多样性及基因组大小比较研究发现, 哺乳动物可能发生过一次或更多次全基因组复制事件, 与Ohno所认为的一致. 由此可见, 全基因组复制事件, 即多倍化可能在脊椎动物的早期演化历程中做出了贡献^[38,81,82]. 在硬骨鱼类多倍化研究中, 1998年, Amores等人^[50]研究斑马鱼(*Danio rerio*)的Hox基因簇时发现, 其拥有的7个Hox基因簇几乎是四足类动物(包括人类、两栖类等)的二倍, 提示在硬骨鱼类进化历程中可能发生过第三次全基因组复制. 随后几年, 研究者基于比较基因组学和同源异型框基因簇的研究, 也提示了辐鳍鱼类在其进化早期, 可能发生过第三次特异的全基因组复制, 即3R假说(the fish-specific genome duplication)^[49,83,84]. 2004年, Vandepoele等人^[38]通过比较人类和红鳍东方鲀的全基因组序列, 推算出第一次和第二次全基因组复制事件大约发生于600 Ma, 而鱼类特异的第三次全基因组复制事件发生的时间大约为320 Ma, 这个结果与Christoffels等人^[84]推测的时间接近, 成为支持3R假说的证据. 但以上研究仅为基于间接证据的推测, 脊椎动物多倍化演化的细节, 未得到明确的阐述.

(2) 硬骨鱼类中的多倍体基因组研究. 辐鳍鱼纲的第三轮特异性全基因组加倍事件可能推进了这一类群

的辐射演化和大量物种形成事件^[49,83,84]. 随后, 多倍化在鲟形目、鲤形目和鲑形目等类群中依然持续发生, 最为引人注目的是鲤形目和鲑形目. 两个类群在后来的演化过程中, 均经历了第四轮特异性的全基因组加倍事件, 某些类群中甚至仍有多倍化现象持续发生.

(i) 鲑科鱼类中的同源多倍体基因组演化. 鲑形目中以鲑科鱼类(*Salmonids*)为代表, 发生了特异性第四轮全基因组加倍事件(*Salmonid-specific fourth vertebrate whole-genome duplication*, Ss4R), 基于减数分裂多价体形成的观察、大量重复基因对来源于两条部分同源染色体等证据提示了其同源多倍化起源^[85,86].

鲑科鱼类中最早报道的是虹鳟(*Oncorhynchus mykiss*)的基因组^[87]. 利用Roche/454测序平台对虹鳟孤雄生殖个体进行测序, 并加测了Solexa/Illumina序列进行纠错, Berthelot等人^[87]获得了虹鳟基因组草图, 并解析了其同源多倍化演化历史. 在90~102 Ma发生Ss4R事件之后, 虹鳟基因组并未发生普遍的染色体重排事件, 其基因组二倍化进程呈现出缓慢和阶段性进行的模式——主要由假基因事件的持续发生导致, 且部分同源基因对之间并未发生大规模的基因间区缩减事件^[87].

通过Sanger和二代测序技术对孤雄发育的大西洋鲑鱼(*Salmo salar*)个体进行测序, 并构建遗传图谱, Lien等人^[88]组装了大西洋鲑鱼基因组草图. 通过对该基因组演化特征的解析, 发现大西洋鲑鱼起源于约80 Ma的同源多倍化事件, 且可能经历了由TE介导的重复序列扩张, 同时造成了全基因组大规模的结构重塑. 该基因组报道了迄今为止脊椎动物中最高比例的重复序列, 其中DNA类型的TE超家族Tc/mariner占比最高(13%), 其扩张与大西洋鲑鱼基因组二倍化事件同时发生. 通过分析TE与大西洋鲑鱼基因组二倍化的关系, 研究者得出结论, Ss4R事件发生后, 在基因丢失事件发生期间, 控制TE活性的相关调控进程失活, 进而导致了TE的扩张; 反过来, TE引发的染色质断裂以及异位重组事件使得全基因组范围发生染色体重排, 进而推进了基因组的二倍化进程^[88,89].

(ii) 鲤科鱼类中多样的多倍化现象以及基因组演化. 近年来, 相关研究对鲤科鱼类系统发育关系有了初步的建立, 并发现多倍体在系统发育树上出现散在分布的情况, 提示不同类群的鱼类可能出现多次基因组加倍的现象^[41], 可能受到杂交和多倍化演化历史的影响, 鲤科鱼类在分类关系的解析上存在诸多争议, 具

体表现为: 基于形态学证据得到的鲤亚科、鲃亚科、裂腹鱼亚科和野鲮亚科等, 在分子系统发育树上却不能形成明确的单系枝; 而基于不同的分子标记可得到不同的系统演化关系. 基于此, 很多学者基于不同的分子系统发育证据, 将这些亚科合并, 并重新分类得到了2~11个系(series)或族(tribes)^[41,90-93]. 除此之外, 基于核基因的研究发现, 这些多倍体物种的多拷贝基因, 在系统发育关系树上呈现了同源加倍(拷贝分歧出现在物种形成之后)或异源加倍(拷贝分歧出现在物种形成之前)的演化历史^[40]. 但该类群鱼类究竟该如何分类、其多倍化演化历史究竟如何, 仍需要选取更多的代表物种, 基于全基因组分子标记来进行系统发育关系的构建, 并有效地将表型组数据化, 综合还原其杂交和全基因组加倍历史, 才能更好地还原这一类群的演化历史.

随着基因组研究的开展, 近年来陆续报道了鲤科鱼类中的多倍体基因组, 其中以鲤鱼和金线鲃基因组为代表^[94,95]. Xu等人^[94]利用二代测序、BAC-end文库以及遗传图谱组装了雌核发育鲤鱼基因组, 并解析了鲤亚科鱼类在硬骨鱼类第三轮全基因组加倍的基础上, 发生的鲤亚科特异性的第四轮异源多倍化事件, 并解析了鲤鱼的两个亚种起源事件以及控制鳞片和皮肤颜色的遗传机制. 其次, 杨君兴团队^[95]利用二代测序技术, 组装了金线鲃属三个物种的基因组草图. 虽然有证据表明, 金线鲃属鱼类在系统发育关系上与鲤亚科鱼类更接近, 且可能与鲫属、鲤属鱼类共享了同一次异源多倍化起源事件^[41], 但遗憾的是, 金线鲃基因组文章并未对该属鱼类的多倍化起源事件进行解析. 但该属鱼类基因组的发表, 为未来综合解析鲤亚科鱼类异源多倍化起源事件提供了数据支持.

(3) 脊椎动物新生多倍化体系中的组学研究. 由于测序及分析等技术难度暂时未能获得高质量的参考基因组, 脊椎动物中的新生多倍化体系的研究仅在转录组学层面开展. 初步研究提示了遗传水平和转录组水平的快速变化. 首先, 遗传水平的研究结果提示, 在杂交子代中存在大量的变异和嵌合基因^[96-98]. 例如, 由刘少军和罗静等人^[98]在鲫鲤杂交体系中发现杂交和多倍体子代中有约10%的基因结构为来源于母本鲫鱼和父本鲤鱼特异性变异位点的嵌合基因, 且部分嵌合基因集中分布在与突变形成和修复相关功能的信号通路中; 在鲫鲂杂交体系中存在15.90 M的点突变, 其中

有5.55%的点突变发生在编码区域^[97]. 其次, 杂交子代中的基因表达规律呈现出类似于植物的“转录组休克”现象, 即在“基因组休克”效应的驱动下, 杂交和多倍体子代中呈现出的整体表达水平以及基因差异表达的快速改变^[97-101]. 例如, 在鲫鲤杂交体系中, 四倍体 F_1 和 F_2 代的整体基因表达模式呈现与亲本以及八倍体子代均不相同的模式, 同时基因表达的差异可能与突变、细胞凋亡程序调控等相关^[98]; 在鲂鲮(团头鲂(♀, $2n=48$)×黄尾密鲮(♂, $2n=48$))杂交体系中, F_1 子代存在二倍体($2n=48$)和三倍体($3n=72$), 通过转录组水平分析共鉴别父母本和两种倍性子代都有表达的基因10057个, 分析发现这两种倍性 F_1 子代基因相对亲本都存在非对称的差异表达, 如三倍杂交子代中, 相较于母本团头鲂有3.73%的基因发生上调和1.92%的基因发生下调, 而相较于父本黄尾密鲮有28.85%的基因发生上调和7.80%的基因发生下调, 并且有18.55%的基因是母本(团头鲂)优势表达, 仅5.84%的基因是父本(黄尾密鲮)优势表达^[99]. 这些结果提示在多倍化初期基因组的融合会对基因组和转录水平造成巨大的影响.

基于以上杂交和多倍化体系的转录组研究结果提示, 在杂交和多倍化的初始阶段, 子代的基因结构、基因表达发生变化, 并伴随着比植物多倍体更加持久和叠加的“基因组休克”效应等综合因素都有可能多倍体个体的存活率低、育性差^[98,102], 上述因素提示了导致脊椎动物中多倍体较少的可能原因. 但要详细地解析脊椎动物多倍化发生初期经历的关键改变和决定其存活的关键因子, 仍需要在遗传、基因表达调控和表观遗传改变等层次提供更多的证据^[96,97].

(4) 两栖类多倍体基因组演化研究. 目前两栖类和爬行类多倍体的基因组研究中, 仅报道了异源四倍体非洲爪蟾的基因组. 在经历了17~18 Ma的异源多倍化事件之后, 四倍体爪蟾发生了与植物类似的染色体结构非对称演化, 且部分同源基因(homeologous genes)的两个拷贝经历了不同的演化命运. 由组蛋白修饰、DNA甲基化以及RNA聚合酶绑定等调控机制, 控制两个基因拷贝的表达, 使得其中一个拷贝维持与二倍体祖先相同的表达模式, 另一个拷贝则发生新功能化/亚功能化^[103]. 以上研究的开展, 为解析脊椎动物古多倍化事件发生后基因组的演化, 以及冗余基因拷贝命运等做出了有意义的提示, 也为脊椎动物多倍体组学研究的开展提供了技术参考.

2.3 基因组冗余度和复杂度的增加为多倍体的研究带来挑战

多倍化造成的最直接的结果就是遗传物质的增加, 遗传物质的冗余给多倍体各个层次研究的开展带来了挑战. 多倍化的早期研究是基于细胞中的DNA含量或染色体组型来判定物种是否发生了基因组加倍^[79,104]; 分子生物学手段普及后, 分子水平的研究也得到了开展, 而多倍体中冗余的基因拷贝也为分子水平的研究带来困难——更大的基因组、更多的等位基因及基因拷贝^[2,105], 使研究者花费更多的时间、精力和研究经费进行大量的基因扩增、克隆、测序和分析. 近年来, 由于测序技术和数据分析的长足进步, 以及测序成本的大幅降低, “组学”时代的来临也为多倍化的研究提供了契机, 但同时, 多倍体造成的基因组冗余也为获得可靠的参考基因组带来了挑战. 在数据分析方面, 在许许多多倍体物种中, 由于存在十分丰富而广泛的重复序列, 甚至因为杂合度而可能存在一定比例的嵌合基因^[98,106], 这些问题均会造成小片段无法得到高效正确的拼装, 并且对拼装结果的完整程度也有影响. 尤其在异源多倍体中, 不同来源的基因组信息还存在一定的差异, 这就给拼装引入歧义, 使得相关算法的进行受到限制. 近年来, 在一些多倍体植物的研究中, 不断有相应的算法和程序得到开发、调整而改善拼装质量^[107,108], 但调整的程度所能容忍的杂合度亦是有限的. 多倍体转录组、甲基化组的分析也因难以识别同源基因、部分同源基因座位而更加复杂. 另外, 由于多倍体物种存在多拷贝及更多的等位基因座位, 使得许多基因编辑工作难以实现, 相关的冗余基因演化命运以及功能验证工作也难以开展^[109].

综上, 多倍体遗传物质冗余为多倍体分子水平、组学水平以及基因功能研究均带来不小的挑战. 多倍化不仅造成实验工作量的增加, 同时也增大了实验操作及数据分析的难度. 获得染色体水平的高质量参考基因组, 可能是突破以上研究困难的关键.

3 脊椎动物多倍化研究中理想模型与组学研究技术

基于已有的植物体系研究, 合适的研究模型是获得多倍化发生初期遗传和表观遗传水平快速变化线索

的关键. 但全基因组加倍造成的数据冗余又为研究工作的开展带来挑战, 尤其是在目前的组学时代下, 数据的复杂度意味着测序和分析的难度增加. 因此, 选择合适的研究技术也尤为重要. 获得一个拼装和注释均较为完善的参考基因组, 可为进一步开展全基因组演化历史重建、新生多倍体表观遗传水平和调控网络分析等提供便利^[110].

3.1 多倍体研究模型的选取

基于植物多倍体类群的研究为脊椎动物多倍化研究的开展提供了理论指导和技术参考. 首先, 在多倍化演化的长期效应追踪方面, 利用发生过或仍在经历全基因组加倍的类群作为研究模型, 可为厘清多倍体类群杂交和全基因组加倍演化历史提供线索. 在脊椎动物中, 鲤科鱼类^[40,41]和非洲爪蟾属^[45]等均符合这些条件. 其次, 基于人工培育的多倍化体系, 则为追踪多倍化发生初期遗传和表观遗传各层次改变提供便利. 例如, 植物中的拟南芥属^[111,112]、棉属^[9,36]和芸薹属^[19,37], 这些存在种间甚至种内倍性多态的类群, 甚至可通过杂交进行人工培育多倍化体系, 都可作为研究植物类群杂交和多倍化基因组演化、以及全基因组加倍后初期时代各层次变化的最佳模型. 在硬骨鱼类中, 鲫属鱼类清晰的异源多倍化演化历史、其多次多倍化的演化特征, 以及其与不同属甚至不同亚科鱼类之间的杂交和基因组加倍耐受力, 均使得该属鱼类及其人工杂交多倍化体系成为追踪脊椎动物杂交和多倍化发生后关键变化的理想模型.

3.2 脊椎动物多倍体基因组研究技术的发展

随着基因组测序技术的不断更新和成本的持续降低, 组学分析可以有效克服多倍体遗传物质冗余带来的分子水平和基因功能研究方面的困难. 要获得高质量的参考基因组, 需要发展合适的测序和组装策略, 以辅助多倍体基因组复杂区域的正确组装.

首先, 在测序策略方面, 在植物多倍化体系中, 已有许多研究通过获得多倍体的加倍单倍型基因组(doubled-haploid genome), 例如通过诱导植株花药染色体组加倍, 发育成植株进行测序^[110]. 由于这样的加倍单倍型基因组, 其复杂度在理论上仅为多倍体基因组的一半, 从而降低了多倍体的遗传杂合度, 使得数据的获得和分析都得到了技术性突破. 在脊椎动物中,

目前已报道可以进行无性生殖的多倍体有蟾蜍和鲤科鱼类(doubled-haploid)等^[42,94,113]。通过获得雌核发育的遗传杂合度降低的个体进行全基因组测序,可为进一步研究多倍化体系初期基因丢失、突变,基因表达,表观遗传水平变异以及调控网络的变化规律提供可靠的参考基因组。

第二,在测序技术整合方面,由于多倍化引发的一系列效应,如重复区域的增加、大量的旁系同源基因、同源多倍体中分歧度较低的部分同源序列、以及异源多倍体中由于杂交而引发的高杂合度等,都可能导致基因组拼装的片段化^[114]。近年来,利用三代测序、BAC文库测序等技术手段,构建长片段(>20 Kb)文库进行测序,辅助大量二代测序数据进行拼装,也是提升多倍体基因组拼装完整度的有效策略。例如, Pacific Bioscience公司开发的单分子实时测序技术(single molecule, real-time sequencing technology, SMRT)有效提升了序列读长。长片段的引入辅助,可以显著提升GC区域、重复区域的拼装完整性,也为Scaffold拼接中的补洞(gap filling)步骤提供便利^[115-117]。其次, BioNano Genomics公司通过获得DNA分子酶切图谱,可得到千碱基甚至百万碱基长度的光学图谱数据,并与组装的scaffold互相延伸,形成超长super scaffold^[118],

辅助SMRT测序技术获得高质量的组装结果,如在山羊基因组中scaffold N50达到了20 Mb^[119]。此外,随着染色质构象捕获测序技术的发展,使得高复杂基因组在缺乏物理图谱或遗传图谱的情况下,都可实现染色体级别的组装,如基因组大小高达4.6 Gb的大麦基因组^[120]。这些新技术的发展,为获得高质量的多倍体物种参考基因组提供了便利。

第三,在基因功能研究方面,多倍体脊椎动物基因编辑相关研究尚未有报道的原因,一方面可能是多拷贝和多等位基因座位带来的技术困难,另一方面,也可能由于缺乏遗传背景解析清楚的参考基因组。Yin等人^[109]基于CRISPR-cas9基因编辑技术提出了异源多倍体脊椎动物冗余基因拷贝演化命运验证的实验设计,以红鲫中可能发生功能分化的*fgf20a*基因的两个拷贝为例,通过设计基因拷贝的敲除,追踪基因拷贝蛋白表达水平,并观察胚胎发育过程中的致死率、出苗率等,以确认该基因两个拷贝的演化命运,以期对未来多倍体脊椎动物的基因功能研究提供参考思路。

总之,合适的测序策略和技术,如加倍单倍型测序、长片段测序技术、光学图谱以及染色质构象捕获测序辅助组装等技术的引入,可能可以有效地克服基因组复杂度带来的困难,并在缺乏物理图谱和遗传图

图 1 脊椎动物多倍体基因组及演化研究策略

Figure 1 The research strategies for genome and evolution in polyploid vertebrates

谱的情况下, 获得组装至染色体水平的基因组。但也存在一些亟待解决的问题, 如长片段测序技术存在的高比例测序错误, 在结合二代测序数据的拼装过程中, 需要一些算法辅助进行纠错并提升运算效率; 多倍体物种由于基因组较大或染色体数目较多, 相应地会产生更多的原始数据, 需要开发高效的拼装程序加以辅助; 除此之外, 大量重复序列及多拷贝基因的正确座位识别等, 也需要相应的软件开发进行完善^[107,108,121~123]。获得高质量的多倍体基因组, 可为全基因组水平还原多倍体类群的起源历史和解析其长期演化效应提供有效的分子标记, 也可为遗传和表观遗传水平追踪多倍化发生初期“休克”效应下的关键因子提供可靠的参考基因组, 同时, 基于可靠的基因组遗传背景分析, 也为多倍体基因功能实验的开展提供依据。

综上, 植物和动物体系中, 有关多倍化发生的驱动力以及初期世代发生的改变规律可能存在许多相似之处, 这些相似的变化有待进一步探究。在植物多倍化体系中运用的研究技术和策略、提出的假设和结论等, 可能为脊椎动物多倍化研究提供许多研究理论上的参考或提示(图1)^[124]。鲤科鱼类中的鲫属复合种, 以及红鲫与鲤鱼构建的异源多倍化体系, 均可作为研究脊椎动物多倍化各层次问题的理想模型。无论是天然多倍体群体与环境适应的关系, 或是多倍体形成初期在遗传和基因表达、表观修饰以及网络调控的即时效应方面, 均可利用这一类群或体系获得有意义的线索。合适的研究模型和有效的研究策略相结合, 可解析新生脊椎动物多倍体发生的变化, 并为解释多倍化现象在脊椎动物中罕见的原因提供证据支持。

参考文献

- Comai L. The advantages and disadvantages of being polyploid. *Nat Rev Genet*, 2005, 6: 836–846
- Otto S P, Whitton J. Polyploid incidence and evolution. *Annu Rev Genet*, 2000, 34: 401–437
- Adams K L. Evolution of duplicate gene expression in polyploid and hybrid plants. *J Heredity*, 2007, 98: 136–141
- Wendel J F. Genome evolution in polyploids. *Plant Mol Biol*, 2000, 42: 225–249
- Cui L, Wall P K, Leebens-Mack J H, et al. Widespread genome duplications throughout the history of flowering plants. *Genome Res*, 2006, 16: 738–749
- Tang H, Wang X, Bowers J E, et al. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res*, 2008, 18: 1944–1954
- Liu D C, Fang H. Study on the role of Chinese *Ae. tauschii* in the evolution of Chinese common wheat landraces (in Chinese). *Southwest China J Agric Sci*, 2002, 16: 32–35 [刘登才, 房洪. 中国节节麦在中国特有小麦系统演化中的作用. *西南农业学报*, 2003, 16: 32–35]
- Parisod C, Christin P A. Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytol*, 2008, 178: 436–447
- Paterson A H, Wendel J F, Gundlach H, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*, 2012, 492: 423–427
- Volff J N. Genome evolution and biodiversity in teleost fish. *Heredity*, 2005, 94: 280–294
- Lokki J, Saura A. Polyploidy in insect evolution. *Basic Life Sci*, 1979, 13: 277–312
- Bogart J P. Evolutionary implications of polyploidy in amphibians and reptiles. *Basic Life Sci*, 1979, 13: 341–378
- Mable B K, Alexandrou M A, Taylor M I. Genome duplication in amphibians and fish: An extended synthesis. *J Zool*, 2011, 284: 151–182
- Rodriguez F, Arkhipova I R. Transposable elements and polyploid evolution in animals. *Curr Opin Genets Dev*, 2018, 49: 115–123
- McClintock B. The significance of responses of the genome to challenge. *Science*, 1984, 226: 792–801
- Zhao N, Zhu B, Li M, et al. Extensive and heritable epigenetic remodeling and genetic stability accompany allohexaploidization of wheat. *Genetics*, 2011, 188: 499–510
- Zhang H, Bian Y, Gou X, et al. Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. *Proc Natl Acad Sci USA*, 2013, 110: 3447–3452
- Bao Y, Xu Q. Extensive reprogramming of cytosine methylation in *Oryza* allotetraploids. *Genes Genom*, 2015, 37: 517–524
- Liu S, Liu Y, Yang X, et al. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun*, 2014, 5: 3930

- 20 Flagel L E, Wendel J F, Udall J A. Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. *BMC Genom*, 2012, 13: 302
- 21 Kirkbride R C, Yu H H, Nah G, et al. An epigenetic role for disrupted paternal gene expression in postzygotic seed abortion in *Arabidopsis* interspecific hybrids. *Mol Plant*, 2015, 8: 1766–1775
- 22 Song Q, Guan X, Chen Z J. Dynamic roles for small RNAs and DNA methylation during ovule and fiber development in allotetraploid cotton. *PLoS Genet*, 2015, 11: e1005724
- 23 Wang M, Yuan D, Tu L, et al. Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New Phytol*, 2015, 207: 1181–1197
- 24 Adams K L, Cronn R, Percifield R, et al. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci USA*, 2003, 100: 4649–4654
- 25 Seijo J G, Kovalsky E I, Chalup L M I, et al. Karyotype stability and genome-specific nucleolar dominance in peanut, its wild 4× ancestor, and a synthetic AABB polyploid. *Crop Sci*, 2018, 58: 1671–1683
- 26 Gao L, Diarso M, Zhang A, et al. Heritable alteration of DNA methylation induced by whole-chromosome aneuploidy in wheat. *New Phytol*, 2016, 209: 364–375
- 27 Thorpe P H, González-Barrera S, Rothstein R. More is not always better: The genetic constraints of polyploidy. *Trends Genets*, 2007, 23: 263–266
- 28 Chen Z J. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol*, 2007, 58: 377–406
- 29 Ni Z, Kim E D, Ha M, et al. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature*, 2009, 457: 327–331
- 30 Lynch M, Conery J S. The evolutionary fate and consequences of duplicate genes. *Science*, 2000, 290: 1151–1155
- 31 Sémon M, Wolfé K H. Consequences of genome duplication. *Curr Opin Genets Dev*, 2007, 17: 505–512
- 32 Mable B K. “Why polyploidy is rarer in animals than in plants”: Myths and mechanisms. *Biol J Linnean Soc*, 2004, 82: 453–466
- 33 Wertheim B, Beukeboom L W, van de Zande L. Polyploidy in animals: Effects of gene expression on sex determination, evolution and ecology. *Cytogenet Genome Res*, 2013, 140: 256–269
- 34 Wang J, Ye L H, Liu Q Z, et al. Rapid genomic DNA changes in allotetraploid fish hybrids. *Heredity*, 2015, 114: 601–609
- 35 Gasser M. Genetic-ecological Investigations in *Biscutella levigata* L. Dissertation for Doctoral Degree. Gallenkirch: Swiss Federal Institute of Technology Zurich, 1986
- 36 Salmon A, Flagel L, Ying B, et al. Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytologist*, 2010, 186: 123–134
- 37 Xu Y, Zhao Q, Mei S, et al. Genomic and transcriptomic alterations following hybridisation and genome doubling in trigeneric allohexaploid *Brassica carinata*×*Brassica rapa*. *Plant Biol*, 2012, 14: 734–744
- 38 Vandepoele K, De Vos W, Taylor J S, et al. Major events in the genome evolution of vertebrates: Paraneome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc Natl Acad Sci USA*, 2004, 101: 1638–1643
- 39 Hardie D C, Hebert P D. Genome-size evolution in fishes. *Can J Fish Aquat Sci*, 2004, 61: 1636–1646
- 40 Ma W, Zhu Z H, Bi X Y, et al. Allopolyploidization is not so simple: Evidence from the origin of the tribe Cyprinini (Teleostei: Cypriniformes). *Curr Mol Med*, 2014, 14: 1331–1338
- 41 Zhang G, Tian Y, Zhang J, et al. Hybrid *de novo* genome assembly of the Chinese herbal plant danshen (*Salvia miltiorrhiza* Bunge). *Gigascience*, 2015, 4: 62
- 42 Stöck M, Ustinova J, Betto-Colliard C, et al. Simultaneous Mendelian and clonal genome transmission in a sexually reproducing, all-triploid vertebrate. *Proc R Soc B-Biol Sci*, 2012, 279: 1293–1299
- 43 Gruber S L, Silva A P Z, Haddad C F B, et al. Cytogenetic analysis of *Phyllomedusa distincta* Lutz, 1950 ($2n = 2x = 26$), *P. tetraploidea* Pombal and Haddad, 1992 ($2n = 4x = 52$), and their natural triploid hybrids ($2n = 3x = 39$) (Anura, Hylidae, Phyllomedusinae). *BMC Genet*, 2013, 14: 75
- 44 Song C, Liu S J, Xiao J, et al. Polyploid organisms. *Sci China Life Sci*, 2012, 55: 301–311
- 45 Kawahara H. Production of triploid and gynogenetic diploid *Xenopus* by cold treatment. *Dev Growth Differ*, 1978, 20: 227–236
- 46 Bogart J P, Bi K. Genetic and genomic interactions of animals with different ploidy levels. *Cytogenet Genome Res*, 2013, 140: 117–136
- 47 Manning G J, Cole C J, Dessauer H C, et al. Hybridization between parthenogenetic lizards (*Aspidoscelis neomexicana*) and gonochoristic

- lizards (*Aspidoscelis sexlineata viridis*) in New Mexico: Ecological, morphological, cytological, and molecular context. *Am Museum Novitates*, 2005, 3492: 1
- 48 Danielayn F, Arakelyan M, and Stepanyan I. The progress of microevolution in hybrids of Rock lizards of genus *Darevskia*. *Biol J Armenia*, 2008, 60: 147-156
- 49 Ravi V, Venkatesh B. Rapidly evolving fish genomes and teleost diversity. *Curr Opin Genets Dev*, 2008, 18: 544-550
- 50 Amores A, Force A, Yan Y L, et al. Zebrafish hox clusters and vertebrate genome evolution. *Science*, 1998, 282: 1711-1714
- 51 Leggatt R A, Iwama G K. Occurrence of polyploidy in the fishes. *Rev Fish Biol Fisheries*, 2003, 13: 237-246
- 52 Taylor J S, Braasch I, Frickey T, et al. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res*, 2000, 13: 382-390
- 53 Phillips R, Ráb P. Chromosome evolution in the Salmonidae (Pisces): An update. *Biol Rev*, 2001, 76: 1-25
- 54 Thorgaard G H and Gall G A E. Adult triploids in a rainbow trout family. *Genetics*, 1979, 93: 961-973
- 55 Arai R, Fish Karyotypes: A Check List. Tokyo: Springer Japan, 2011
- 56 Froese R, Pauly D, eds. FishBase. www.fishbase.org version (1/2015). 2015
- 57 Jose A. Carmona, Oris I. Sanjur, Ignacio Doadrio, et al. Hybridogenetic reproduction and maternal ancestry of polyploid iberian fish: The *Tropidophoxinellus alburnoides* complex. *Genetics*, 1997, 146: 983-993
- 58 Wu Y F, Kang B, Men Q, et al. Chromosome diversity of Tibetan fishes (in Chinese). *Zool Res*. 1999, 20: 258-264 [武云飞, 康斌, 门强, 等. 西藏鱼类染色体多样性的研究. *动物学研究*, 1999, 20: 258-264]
- 59 Luo J, Gao Y, Ma W, et al. Tempo and mode of recurrent polyploidization in the *Carassius auratus* species complex (*Cypriniformes*, *Cyprinidae*). *Heredity*, 2014, 112: 415-427
- 60 Liu X L, Jiang F F, Wang Z W, et al. Wider geographic distribution and higher diversity of hexaploids than tetraploids in *Carassius* species complex reveal recurrent polyploidy effects on adaptive evolution. *Sci Rep*, 2017, 7: 5395
- 61 Ren L, Gao X, Yang C, et al. Comparison of diploid and triploid *Carassius auratus* provides insights into adaptation to environmental change. *Sci China Life Sci*, 2018, 61: 1407-1419
- 62 Liu S, Liu Y, Zhou G, et al. The formation of tetraploid stocks of red crucian carp×common carp hybrids as an effect of interspecific hybridization. *Aquaculture*, 2001, 192: 171-186
- 63 Liu S, Qin Q, Xiao J, et al. The formation of the polyploid hybrids from different subfamily fish crossings and its evolutionary significance. *Genetics*, 2007, 176: 1023-1034
- 64 Tao H F. Studies on the biological characteristics of the distant hybrids between koi carp and flower crucian carp (in Chinese). Dissertation for Master's Degree. Changsha: Hunan Normal University, 2011 [陶穗菲. 锦鲤和花鲫远缘杂交后代生物学特性研究. 硕士学位论文. 长沙: 湖南师范大学, 2011]
- 65 Liu S, Qin Q, Wang Y, et al. Evidence for the formation of the male gynogenetic fish. *Mar Biotech*, 2010, 12: 160-172
- 66 Liu S J. Fish Distant Hybridization (in Chinese). Beijing: Science Press, 2015. 1-274 [刘少军. 鱼类远缘杂交. 北京: 科学出版社, 2015. 1-274]
- 67 He W, Qin Q, Liu S, et al. Organization and variation analysis of 5S rDNA in different ploidy-level hybrids of red crucian carp×topmouth culter. *PLoS ONE*, 2012, 7: e38976
- 68 Xiao J. Establishment of hybrid strains between blunt snout bream and topmouth culter and their genetic characteristic research (in Chinese). Dissertation for Master's Degree. Changsha: Hunan Normal University, 2013 [肖军. 异源鲂鲢杂交品系的建立及其遗传特性研究. 硕士学位论文. 长沙: 湖南师范大学. 2013]
- 69 Zhang Z H, Chen J, Li L, et al. Research advances in animal distant hybridization. *Sci China Life Sci*, 2014, 57: 889-902
- 70 Hu J, Liu S, Xiao J, et al. Characteristics of diploid and triploid hybrids derived from female *Megalobrama amblycephala* Yih×male *Xenocypris davidi* Bleeker. *Aquaculture*, 2012, 364-365: 157-164
- 71 He W, Xie L, Li T, et al. The formation of diploid and triploid hybrids of female grass carp×male blunt snout bream and their 5S rDNA analysis. *BMC Genet*, 2013, 14: 110
- 72 Yu J, Hu S, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 2002, 296: 79-92
- 73 Goff S A, Ricke D, Lan T H, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 2002, 296: 92-100
- 74 Zhang T, Hu Y, Jiang W, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotech*, 2015, 33: 531-537

- 75 Li F, Fan G, Lu C, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotech*, 2015, 33: 524–530
- 76 Yang J, Liu D, Wang X, et al. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat Genet*, 2016, 48: 1225–1232
- 77 Hegarty M J, Hiscock S J. Genomic clues to the evolutionary success of polyploid plants. *Curr Biol*, 2008, 18: R435–R444
- 78 Wang J, Tian L, Madlung A, et al. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics*, 2004, 167: 1961–1973
- 79 Ohno S. Evolution by gene duplication. London: George Allen & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag. 1970
- 80 Furlong R F, Holland P W H. Polyploidy in vertebrate ancestry: Ohno and beyond. *Biol J Linnean Soc*, 2004, 82: 425–430
- 81 Dehal P, Boore J L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol*, 2005, 3: e314
- 82 Donoghue P C J, Purnell M A. Genome duplication, extinction and vertebrate evolution. *Trends Ecol Evol*, 2005, 20: 312–319
- 83 Meyer A, Van de Peer Y. From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). *Bioessays*, 2005, 27: 937–945
- 84 Christoffels A, Koh E G L, Chia J M, et al. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol*, 2004, 21: 1146–1151
- 85 Allendorf F W, Danzmann R G. Secondary tetrasomic segregation of MDH-B and preferential pairing of homeologues in rainbow trout. *Genetics*, 1997, 145: 1083–1092
- 86 Allendorf F W, Thorgaard G H. Tetraploidy and the evolution of salmonid fishes. In: Turner J B, ed. Evolutionary Genetics of Fishes. New York: Plenum Press Corp, 1984. 1–53
- 87 Berthelot C, Brunet F, Chalopin D, et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun*, 2014, 5: 3657
- 88 Lien S, Koop B F, Sandve S R, et al. The Atlantic salmon genome provides insights into rediploidization. *Nature*, 2016, 533: 200–205
- 89 de Boer J G, Yazawa R, Davidson W S, et al. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genom*, 2007, 8: 422
- 90 Zardoya R, Doadrio I. Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *J Mol Evol*, 1999, 49: 227–237
- 91 Gilles A, Lecointre G, Miquelis A, et al. Partial combination applied to phylogeny of European cyprinids using the mitochondrial control region. *Mol Phylogenets Evol*, 2001, 19: 22–33
- 92 Cunha C, Mesquita N, Dowling T E, et al. Phylogenetic relationships of Eurasian and American cyprinids using cytochrome b sequences. *J Fish Biol*, 2002, 61: 929–944
- 93 Liu H, Chen Y. Phylogeny of the East Asian cyprinids inferred from sequences of the mitochondrial DNA control region. *Can J Zool*, 2003, 81: 1938–1946
- 94 Xu P, Zhang X, Wang X, et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat Genet*, 2014, 46: 1212–1219
- 95 Yang J, Chen X, Bai J, et al. The Sinocyclocheilus cavefish genome provides insights into cave adaptation. *BMC Biol*, 2016, 14: 1
- 96 Liu Q, Qi Y, Liang Q, et al. The chimeric genes in the hybrid lineage of *Carassius auratus cuvieri* (♀) × *Carassius auratus red var.* (♂). *Sci China Life Sci*, 2018, 61: 1079–1089
- 97 Qin Q, Cao L, Wang Y, et al. Correction to: Rapid genomic and genetic changes in the first generation of autotetraploid lineages derived from distant hybridization of *Carassius auratus Red Var.* (♀) × *Megalobrama amblycephala* (♂). *Mar Biotech*, 2019, doi: 10.1007/s10126-018-09872-9
- 98 Liu S, Luo J, Chai J, et al. Genomic incompatibilities in the diploid and tetraploid offspring of the goldfish × common carp cross. *Proc Natl Acad Sci USA*, 2016, 113: 1327–1332
- 99 Li W, Liu J, Tan H, et al. Asymmetric expression patterns reveal a strong maternal effect and dosage compensation in polyploid hybrid fish. *BMC Genom*, 2018, 19: 517
- 100 Ren L, Li W, Tao M, et al. Homoeologue expression insights into the basis of growth heterosis at the intersection of ploidy and hybridity in Cyprinidae. *Sci Rep*, 2016, 6: 27040
- 101 Li W, Tan H, Liu J, et al. Comparative analysis of testis transcriptomes associated with male infertility in triploid cyprinid fish. *Reprod Fertil Dev*, 2019, 31: 248–260

- 102 Xu K, Wen M, Duan W, et al. Comparative analysis of testis transcriptomes from triploid and fertile diploid cyprinid fish. *Biol Reprod*, 2015, 92: 95
- 103 Session A M, Uno Y, Kwon T, et al. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*, 2016, 538: 336–343
- 104 Nei M, Nozawa M. Roles of mutation and selection in speciation: From Hugo de Vries to the modern genomic era. *Genome Biol Evol*, 2011, 3: 812–829
- 105 Otto S P. The evolutionary consequences of polyploidy. *Cell*, 2007, 131: 452–462
- 106 Flagel L E, Wendel J F. Gene duplication and evolutionary novelty in plants. *New Phytol*, 2009, 183: 557–564
- 107 Sankoff D, Zheng C, Zhu Q. Polyploids, genome halving and phylogeny. *Bioinformatics*, 2007, 23: i433–i439
- 108 Abbott J C, Butcher S A. Strategies towards sequencing complex crop genomes. *Genome Biol*, 2012, 13: 322
- 109 Yin F Q, Liu W F, Chai J, et al. CRISPR/Cas9 application for gene copy fate survey of polyploid vertebrates. *Front Genet*, 2018, 9: 260
- 110 Feuillet C, Leach J E, Rogers J, et al. Crop genome sequencing: Lessons and rationales. *Trends Plant Sci*, 2011, 16: 77–88
- 111 Ng D W K, Miller M, Yu H H, et al. A role for CHH methylation in the parent-of-origin effect on altered circadian rhythms and biomass heterosis in *Arabidopsis* intraspecific hybrids. *Plant Cell*, 2014, 26: 2430–2440
- 112 Miller M, Zhang C, Chen Z J. Ploidy and hybridity effects on growth vigor and gene expression in *Arabidopsis thaliana* hybrids and their parents. *G3*, 2012, 2: 505–513
- 113 Wang Y, Lu Y, Zhang Y, et al. The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nat Genet*, 2015, 47: 625–631
- 114 Saski C A, Feltus F A, Parida L, et al. BAC sequencing using pooled methods. *Methods Mol Biol*, 2015, 1227: 55–67
- 115 Shin S C, Ahn D H, Kim S J, et al. Advantages of single-molecule real-time sequencing in high-GC content genomes. *PLoS ONE*, 2013, 8: e68824
- 116 Gupta P K. Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotech*, 2008, 26: 602–611
- 117 McCarthy A. Third generation DNA sequencing: Pacific biosciences' single molecule real time technology. *Chem Biol*, 2010, 17: 675–676
- 118 Staňková H, Hastie A R, Chan S, et al. BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. *Plant Biotech J*, 2016, 14: 1523–1531
- 119 Bickhart D M, Rosen B D, Koren S, et al. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat Genet*, 2017, 49: 643–650
- 120 Mascher M, Gundlach H, Himmelbach A, et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature*, 2017, 544: 427–433
- 121 Chin F Y L, Leung H C M, Yiu S M. Sequence assembly using next generation sequencing data—challenges and solutions. *Sci China Life Sci*, 2014, 57: 1140–1148
- 122 Chin C S, Alexander D H, Marks P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*, 2013, 10: 563–569
- 123 Gruenheit N, Deusch O, Esser C, et al. Cutoffs and k-mers: Implications from a transcriptome study in allopolyploid plants. *BMC Genom*, 2012, 13: 92
- 124 Chai J, Su Y, Huang F, et al. The gap in research on polyploidization between plants and vertebrates: Model systems and strategic challenges. *Sci Bull*, 2015, 60: 1471–1478

Genomic evolution in polyploid vertebrates: Models and progress in research techniques

CHAI Jing¹, LIN GuoLiang¹, HUANG Feng¹ & ROBERT W. Murphy^{2,3}

1 State Key Laboratory for Conservation and Utilization of Bio-resource and School of Life Sciences, Yunnan University, Kunming 650091, China;

2 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming 650223, China;

3 Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto M5S 2C6, Ontario, Canada

Polyploidization, also whole-genome duplication (WGD), is ubiquitous in plants but in vertebrates it occurs relatively rarely. The mechanisms that allow polyploid plants to overcome “genome shock” and achieve genomic stability, and how doubled genomes evolve had been well characterized. These studies yield several models that serve as theoretical references for analyses of polyploid vertebrates. Herein, we review the development of theories in WGD studies, summarize the research models and progress in vertebrates, and discuss the challenges of “omics” research and effective technical strategies. We hope to discover feasible means of unravelling the genome evolution of polyploid vertebrates.

vertebrates, polyploidization, omics, research mode

doi: [10.1360/N052018-00210](https://engine.scichina.com/doi/10.1360/N052018-00210)