

植物古基因组学研究进展

张太奎^{1,2}, 苑兆和^{1,2}

1. 南京林业大学, 南方现代林业协同创新中心, 南京 210037

2. 南京林业大学林学院, 南京 210037

摘要: 植物古基因组学是基因组学一个新兴分支, 从现存物种中重建其祖先基因组, 推断在古历史中导致形成现存物种的进化或物种形成事件。高通量测序技术的不断革新使测序读长更长、更准确, 加快了植物参考基因组序列的组装进程, 为古基因组学研究提供了大批量可靠的现存物种的基因组序列资源。全基因组复制 (whole-genome duplication, WGD) 亦称古多倍化, 使植物基因组快速重组, 丢失大量基因, 增加结构变异, 对植物进化极其重要。本文综述了植物基因组测序与组装研究进展、植物古基因组学的原理、植物基因组 WGD 事件以及植物祖先基因组进化场景, 并对未来植物古基因组学研究进行了展望。

关键词: 植物基因组; 测序与组装; 古基因组学; 全基因组复制; 多倍化

Progress in plant paleogenomics

Taikui Zhang^{1,2}, Zhaohe Yuan^{1,2}

1. Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China;

2. College of Forestry, Nanjing Forestry University, Nanjing 210037, China

Abstract: As a new branch of genomics, plant paleogenomics reconstructs ancestral genomes from actual modern species and infers palaeohistory, evolutionary and/or speciation events that have shaped the modern species. Advances in high-throughput sequencing technologies yield accurate long reads, promote the progress of plant genome sequence assembly, and thereby offer paleogenomics a large collection of valuable reference genomes from modern species. Whole-genome duplication (WGD) and polyploidization cause rapid genomic reorganization, massive gene losses and structural variations. WGD events are therefore central to plant evolution. In this review, we summarize recent progress in sequencing and assembly of plant genomes, principles of plant paleogenomics, WGD events in plant genomes, and the most likely evolutionary scenario in plants. Furthermore, we highlight some of the challenges as well as future directions.

Keywords: plant genome; sequence and assembly; paleogenomics; whole-genome duplication; polyploidization

收稿日期: 2017-05-25; 修回日期: 2017-09-04

基金项目: 南京林业大学高层次人才科研启动基金项目(编号: GXL2014070), 南京林业大学优秀博士学位论文创新基金项目, 江苏省普通高校研究生科研创新计划项目(编号: KYLX16_0857)和江苏省高校优势学科建设工程项目(PAPD)资助[Supported by the Initiative Project for Talents of Nanjing Forestry University (No. GXL2014070), the Doctorate Fellowship Foundation of Nanjing Forestry University, the Research Fund for Postgraduate Innovation Project of Jiangsu Province (No. KYLX16_0857), and the Priority Academic Program Development of Jiangsu High Education Institutions (PAPD)]

作者简介: 张太奎, 博士研究生, 研究方向: 植物基因组学。E-mail: taikuizhang@126.com

通讯作者: 苑兆和, 博士, 教授, 研究方向: 经济林培育与分子生物学。E-mail: zhyuan88@hotmail.com

DOI: 10.16288/j.ycz.17-191

网络出版时间: 2017/12/22 11:34:01

URI: <http://kns.cnki.net/kcms/detail/11.1913.R.20171222.1133.001.html>

植物古基因组学是一门从现存物种中重建和分析其祖先基因组的学科,旨在研究植物祖先基因组的进化历史^[1,2]。随着测序技术不断革新,测序成本不断降低,目前已公布了 180 多种植物的参考基因组数据^[3],其中甜橙(*Citrus sinensis*)^[4]、黄瓜(*Cucumis sativus*)^[5]和巨桉(*Eucalyptus grandis*)^[6]等植物基因组组装到染色体水平;海草(*Zostera marina*)^[7]、挪威云杉(*Picea abies*)^[8]和毛竹(*Phyllostachys heterocycla*)^[9]等植物基因组组装到 Scaffold 水平,这为古基因组学研究提供了大量有价值的序列数据资源。古基因组学计算依赖于序列测序准确度和组装水平,低质量的序列会影响古基因组学计算的准确度。如何提高基因组测序和组装的准确度和长度一直是本研究领域所关注的热点。

近年来,大批量植物基因组测序和重测序项目推动了古基因组学的发展,提出的全基因组复制(whole-genome duplication, WGD)(或古多倍化)事件是植物基因产生新功能化的主要进化推动力^[10,11],为植物适应基因重复提供原始材料,有益于植物适应生理和遗传改变^[12]。有研究报道,通过古基因组学研究可溯源到由 7931 条原基因模型和 7 条染色体构成的双子叶植物祖先基因组,以及由 9138 条原基因模型和 5 条染色体构成的单子叶植物祖先基因组^[1]。而在裸子植物进化史中,WGD 事件使松柏类植物与其他裸子植物分化开来^[13],银杏除与被子植物共享一次 WGD 事件外,近期发生过一次银杏特异性 WGD 事件^[14]。由于裸子植物基因组数据有限,裸子植物祖先基因组基因模型尚未确定。

参考基因组序列是古基因组学研究的重要资源,植物全基因组复制事件鉴定是古基因组学研究的重要途径之一。本文主要综述了植物基因组测序与组装研究进展、植物古基因组学的原理、WGD 事件研究历程和鉴定方法及其对植物基因组进化的作用、植物祖先基因组进化场景,并对未来古基因组学研究进行了展望。

1 植物基因组测序与组装

基因组(genome)是指一个物种单倍体的染色体数目及其所携带的全部基因,包括每个染色体的序

列加上细胞器中的任何 DNA,本文主要指植物核基因组。植物基因组通常具有重复序列多、多倍化和杂合度高等特征,使基因组组装复杂化^[15,16]。由于 Sanger、454、SOLid 和 Illumina 测序技术读长短,不能有效地组装基因组复杂重复区域,特别是非自交或重排杂合基因组^[17]。随着高通量测序技术的发展,第三代单分子实时测序(single-molecule real-time sequencing, PacBio)技术使基因组测序读长更长,可以辅助组装出高质量的植物基因组^[3]。近年来,已综合运用 PacBio 和其他测序技术成功组装了凤梨(*Ananas comosus*)、欧洲白桦(*Betula pendula*)、芥菜(*Brassica juncea*)、木豆(*Cajanus cajan*)、藜麦(*Chenopodium quinoa*)、一年生辣椒(*Capsicum annuum*)、中华辣椒(*C. chinense*)、复活草(*Oropetium thomaeum*)、向日葵(*Helianthus annuus*)、大麦(*Hordeum vulgare*)、报春花(*Primula veris*)、海带(*Saccharina japonica*)、丹参(*Salvia miltiorrhiza*)和丝叶狸藻(*Utricularia gibba*)等植物基因组(表 1)。

目前,植物基因组测序项目多采用联合使用多种测序技术的策略,相比单一测序途径具有多种优势。主要表现在:(1)克服高杂合度。例如,凤梨属于自交不亲和、杂合度较高的经济作物。凤梨栽培种‘F153’(*A. comosus* ‘F153’)基因组存在较高杂合度(1%~2%),发现其 k-mer 深度频率统计分布中存在 2 个明显的峰:位于约 110 层深度处的纯合峰和位于约 220 层深度处的杂合峰^[18]。为克服高杂合度,对 F153 和其近缘种(*A. bracteatus*)杂交 F₁ 代进行测序,并综合运用 454、Illumina 和 PacBio 测序技术,组装出高质量(Scaffold N50 高达 11.8 Mb)凤梨基因组;(2)克服复杂倍性。例如,芥菜是 *Brassica* 属异源多倍体作物,基因组 k-mer 频率分布图左峰略微突起,表明其杂合度较低,但复杂倍性使其基因组组装更加困难^[19]。Yang 等^[19]综合运用 Illumina 和 PacBio 技术对芥菜基因组进行组装,基因组 Scaffold N50 高达 855 kb;(3)克服高比例重复序列。例如,复活草基因组是已知草类最小的基因组,其重复序列比例显著高于其他草类物种^[15]。VanBuren 等^[15]结合 Illumina 和 PacBio 测序技术组装得到复活草基因组, Scaffold N50 高达 7.1 Mb,组装效果好。复活草基因组较低的杂合度(0.087%)可能与基因组的高连续

表 1 部分植物基因组组装情况

Table 1 A summary of plant genomes that have been sequenced and assembled

物种	测序技术	基因组大小	重复序列比例(%)	倍性	杂合度(%)	N50 (kb) Scaffold/Contig	文献
凤梨	454+Illumina Hiseq+PacBio	382 Mb	44	二倍体	1~2	11800/126.5	[18]
芥菜	Illumina Hiseq+PacBio	955 Mb	33	异源四倍体	—	855/61.3	[19]
花生基因组 A 亚组	Sanger+454+Illumina Hiseq+ PacBio	1.25 Gb	61	二倍体	—	948/222.93	[22]
花生基因组 B 亚组		1.56 Gb	68			5343.3/234.92	
巨桉	Sanger+Illumina	640 Mb	50	二倍体	—	5000/2261	[6]
一年生辣椒	Sanger+454+Illumina Hiseq+ Illumina GA	3.06 Gb	76	二倍体	0.005	2470/30.0	[23]
中华辣椒			79				
木豆	Sanger+Illumina Hi-seq+Illumina GA	605.78 Mb	51	二倍体	—	516/21.9	[24]
丝叶狸藻	PacBio	101.95 Mb	58	二倍体	—	3424.836	[25]
欧洲白桦	454+SOLid+Illumina Mi-Seq+PacBio	440 Mb	49.23	二倍体	—	527.7	[26]
藜麦	Illumina Hiseq+PacBio	1.39 Gb	64	二倍体	—	3846.917	[27]
大麦	454+Illumina Hiseq+PacBio	4.79 Gb	80.8	二倍体	—	1900/79	[21]
向日葵	Sanger+PacBio	3.6 Gb	41.2	二倍体	—	13.7	[20]
复活草	Illumina Hiseq+PacBio	245 Mb	43	二倍体	0.087	7100/2400	[15]
报春花	Illumina Hiseq+Illumina Miseq+ PacBio	301.8 Mb	7	二倍体	—	164/9.5	[28]
海带	454+Illumina Hiseq+PacBio	537 Mb	40	二倍体	—	252/58.9	[29]
丹参	454+Illumina Hiseq+PacBio	558 Mb	54	二倍体	0.003	51/12.4	[30]

性相关。向日葵基因组 3/4 以上的序列是重复序列，组装难度大^[20]。Badouin 等^[20]使用 PacBio 技术成功组装出 3.6 Gb 向日葵参考基因组序列。大麦基因组较大(4.79 Gb)，而且 80.8%的序列属于重复序列，组装难度大，联合使用 454、Illumina Hiseq 和 PacBio 测序技术提高了 Scaffold 长度和精准度，Scaffold N50 长达 1.9 Mb^[21]；(4)克服祖先基因组倍性复杂且基因组较大的问题。例如，栽培花生(*Arachis hypogaea*)是一个异源四倍体(AABB)，由二倍体花生 *A. duranensis* (AA)和 *A. ipaensis* (BB)杂交及多倍化而来。栽培花生基因组较大(约 2.7 Gb)，重复序列比例高达 64%，基因组组装难度大。经多倍化后，栽培花生亚基因组 A(1.25 Gb)和亚基因组 B(1.56 Gb)改变较小，可以通过两个亚基因组组装成完整基因组，运用多种测序方法对 *A. duranensis* 和 *A. ipaensis* 基因组进行测序，构建了高质量的栽培花生祖先基

因组草图^[22]。对凤梨、芥菜、大麦、向日葵、复活草和花生等基因组的组装表明，联合应用多种测序技术有助于提高基因组的组装效果，提供更可靠的参考基因组序列。

2 植物古基因组学

古基因组学(paleogenomics)最有可能的进化场景主要基于以下假想推断^[31]：(1)基于现存物种间直系同源基因重复等事件追溯祖先基因组；(2)基于插入、删除、融合、分裂、易位等事件的最小数量推断从祖先基因组到现存核型的进化历史。其中 AL(比对长度 $AL = \sum \text{高分基因对长度}$)、CIP($CIP = \sum (\text{高分基因对同源率}/AL) \times 100$)和 CALP($CALP = AL/\text{长度}_{\text{CDS}}$)是古基因组学计算所需的关键参数^[31]。目前，在基于进化关系的比较基因组学中广泛使用

经典的 CIP/CALP 阈值, 该阈值等于 70% 则代表亲缘关系较近的物种基因组间共同祖先可以追溯到 50 百万年前(million year ago, MYA)以内, 该阈值等于 50% 则代表分化较远的物种基因组间共同祖先可以追溯到 50 MYA 以外。Salse 等^[2]提出“三步法”构建祖先基因组, 即分析基因组内共线性, 鉴定系列特异性基因组内改变事件, 推断共有基因组改变事件。不断更新的植物参考基因组序列数据为古基因组学研究提供了更多资源, Murat 等^[32]对显花植物祖先基因组进化模式进行分析, 构建出由 7010 条原基因模型和 7 条染色体构成的草类祖先核型、由 6284 条原基因模型和 7 条染色体构成的双子叶植物祖先核型、由 6707 条原基因模型和 5 条染色体构成的单子叶植物祖先核型以及由 1175 条原基因模型和 15 条染色体构成的显花植物最近共同祖先核型, 首次系统推断被子植物祖先基因组的进化史。

3 植物全基因组复制事件

3.1 WGD 研究历程

研究植物古基因组学的关键途径是确定 WGD 事件。1970 年, Ohno^[33]首次提出多倍化即全基因组复制的假想, 阐释二倍体基因组通过 WGD 进化为四倍体的途径。1997 年, Wolfe 等^[34]在酿酒酵母(*Saccharomyces cerevisiae*)基因组中证实了 WGD 假想。2000 年, 首例植物全基因组草图测序完成, 解析了拟南芥(*Arabidopsis thaliana*)基因组进化历史, 提出拟南芥可能有一个四倍化祖先^[35]。2007 年, 葡萄(*Vitis vinifera*)基因组测序完成, 古基因组学研究表明葡萄的祖先基因组可能是古六倍体^[36]。对葡萄基因组重复事件的研究奠定了双子叶植物祖先基因组共享六倍化的理论基础, 并开启了被子植物祖先基因组进化历史研究的大门。2008 年, 小立碗藓(*Physcomitrella patens*)基因组测序完成。通过古基因组学研究发现, 陆地植物最近共同祖先基因组在多倍化后发生重要基因功能分化, 如(1)丢失适应水生环境、动力蛋白运输相关基因; (2)获得转运能力、信号转导、抗非生物胁迫相关基因; (3)整体提高基因家族复杂度, 这为理解陆地植物进化提供了可靠资

源^[37]。2011 年, 卷柏(*Selaginella moellendorffii*)基因组测序完成, 虽然通过比较基因组学没有发现古多倍化证据, 但为理解陆地植物基因组进化提供了重要资源^[38]。2013 年, 无油樟(*Amborella trichopoda*)基因组测序草图公布, WGD 分析表明其分化发生在古六倍化前, 这为理解被子植物基因组和基因进化提供了重要参考^[39]。2015 年, 凤梨基因组测序完成, 古基因组学研究表明其基因组发生了 2 次 WGD 事件, 染色体重排少, 并且保留了单子叶植物祖先基因组多倍化后 28 条染色体中的 25 条核型, 是保守的单子叶植物参考基因组^[18]。2016 年, 银杏基因组草图完成, 古基因组学研究表明银杏发生了 2 次 WGD 事件, 其中较近 1 次是银杏特异性 WGD^[14]。2017 年, 向日葵基因组测序完成, 古基因组学研究表明其经历了 3 次 WGD 事件: 29 MYA 发生的太阳花种系特异性 WGD, 38~50 MYA 菊分支(Asterids) II 物种共有的 WGD, 以及 122~164 MYA 双子叶植物共有的古六倍化, 为研究菊分支物种古基因组学进化场景提供了依据^[20]。越来越多的植物基因组测序完成, 这为古基因组学研究提供了大量资源(图 1)。显花植物进化研究表明(图 1), 被子植物基因组经历了二倍化(α)、四倍化(β)和六倍化(γ)3 次古 WGD 事件, 禾本目谷类植物基因组经历了 ρ 、 τ 和 σ 3 次 WGD 事件^[40], 植物 ρ 事件发生在禾本目(Poales)内 95~115 MYA, 小麦、玉米和水稻分化之前, 草类和凤梨分化之后^[18]。

3.2 WGD 鉴定方法

目前, 主要通过基因组共线性分析、旁系同源基因 Ks 峰值分布和祖先基因系统发育 3 种证据链鉴定基因组 WGD 事件。通过基因组共线性分析发现基因组内存在大量共线区域, 是古 WGD 事件存在的直接证据^[18,41], 可以通过 SynMap(<https://genomeevolution.org/coge/SynMap.pl>)鉴定基因组内部共线性区域及深度和基因组间共线性区域^[42]。例如, 从凤梨基因组内部鉴定出的 388 个共线性区域包含 64% 注释的基因, 并分布在 25 个连锁群, 表明凤梨基因组进化历史中存在 WGD 事件; 共线深度分析发现 35% 基因组存在不止 1 处重复区域, 说明在凤梨系谱中发生了多次 WGD 事件^[18]。在基因组间共线性分析中,

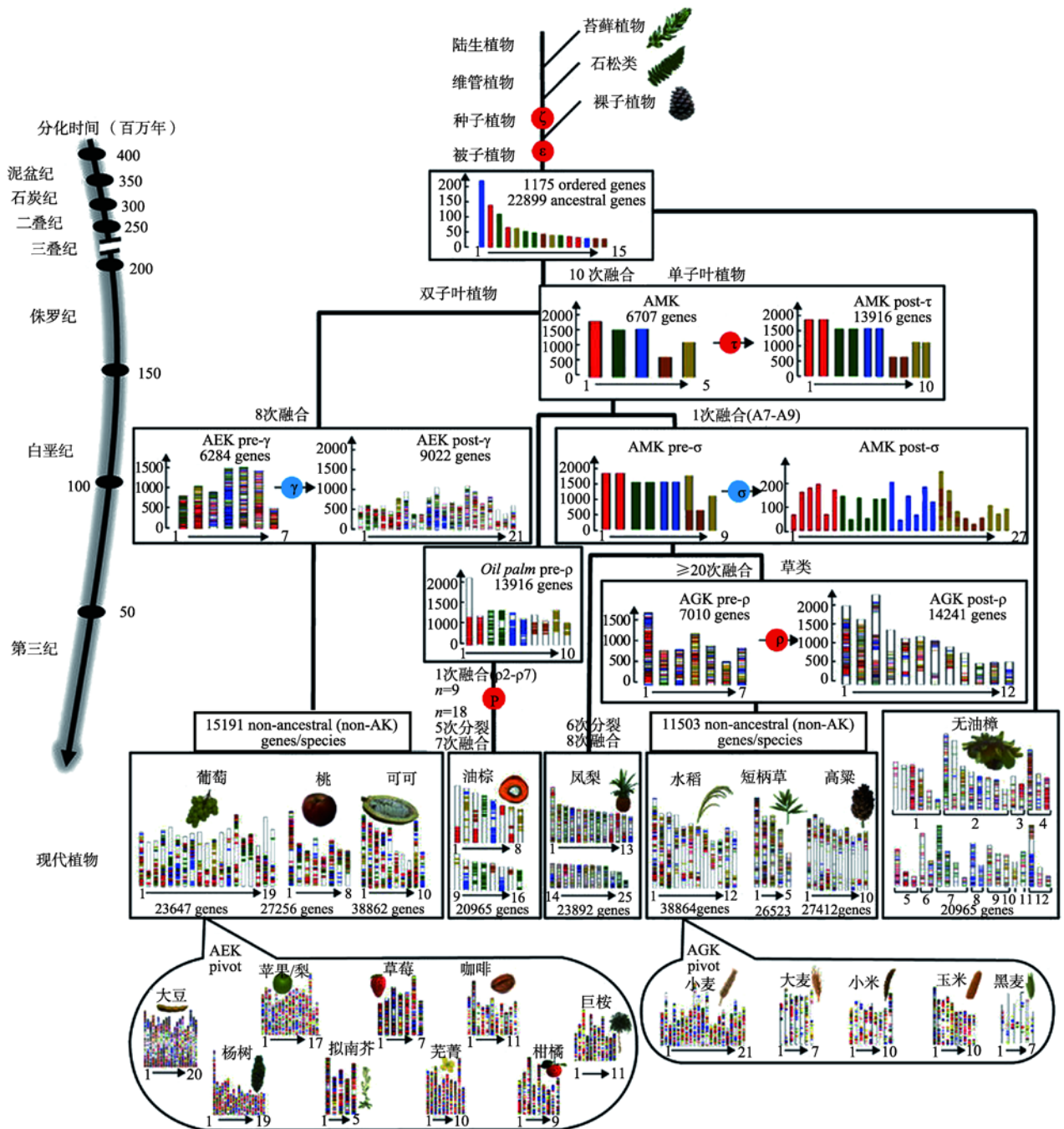


图 1 被子植物古基因组进化场景

Fig. 1 A paleohistorical scenario of angiosperm genomes

参考文献[32]并修改。

无油樟 葡萄(1-3)共线性区域定位到古六倍化事件,说明检测到的WGD事件是2种物种共有的。基因组旁系同源基因 K_s 分布分析是鉴定WGD事件的一种常用途径^[14,43],先通过Orthomcl聚类获取旁系同源基因对,然后依据 K_s 计算值绘制分布图。

WGD会导致基因重复峰值的产生,最后通过识别峰值确定WGD事件。但是,串联重复基因积累会影响 K_s 峰值分布。通过旁系同源基因 K_s 分布分析已成功在毛竹^[9]、鹰嘴豆(*Cicer arietinum*)^[44]、大豆(*Glycine max*)^[45]、小兰屿蝴蝶兰(*Phalaenopsis eq-*

uestris)^[46]和银杏^[14]等物种基因组中鉴定出 WGD 事件。祖先基因系统发育分析也是一种鉴定 WGD 事件的有效途径,即通过系统发育分析识别祖先重复节点基因,进而推断分化时间,再通过祖先基因组重复基因分化时间分布峰值图判断物种间 WGD 事件^[39,46]。通过系统发育研究得到 414 个无油樟和其他被子植物祖先基因共有节点,其中 62% 节点构成第一个峰,发生在 244 MYA, 38% 节点构成第二个峰,发生在 341 MYA^[39]。

3.3 WGD 在植物进化中的作用

WGD 在植物基因组塑形方面具有重要作用,伴随基因丢失的 WGD 事件被视为植物基因新功能化的主要进化力量^[10,11]。泥炭藓纲(Sphagnopsida)最近 WGD 事件早于泥炭藓属(*Sphagnum*)从泥炭藓纲其他两个属分化出来的时间,表明最近 WGD 已成为优先于泥炭藓分化的重要因子,并促进了它们在泥炭湿地中的生态主导地位^[47]。在核心真双子叶植物基因组 γ 事件后,核心真双子叶植物分支物种中祖先基因组散存的重复基因起源于大规模基因搬迁,表明 γ 事件之后的大规模基因搬迁与核心真双子叶植物分支多样化相关^[48]。

此外,WGD 还为植物适应基因重复的形成提供了原始材料,多倍体比二倍体表现出更强的离子吸收能力和抗逆能力,这种生理关联有益于植物适应新环境或具有挑战性的环境,同等条件下多倍体还会增加种群内适应的速度^[12]。同源多倍体来源于种内 WGD 事件,异源多倍体来源于种间杂交^[49]。芸薹属(*Brassica*)植物经历了 3 次基因组 WGD 事件(α 、 β 和 γ)^[50,51]和种系全基因组三倍化事件^[19],是基因组多倍化研究的重要模式物种。海岸红杉(*Sequoia sempervirens*)是裸子植物中罕见的同源六倍体,是研究裸子植物多倍化的重要材料。海岸红杉基因组 WGD 事件表明:裸子植物稀有的多倍化可能是由该分支缓慢的二倍化导致的^[52]。

4 植物祖先基因组进化

图 1 描绘了被子植物祖先基因组的进化历程,双子叶植物祖先基因组有 7 条染色体,单子叶植物

祖先基因组有 5 条染色体,草类祖先基因组有 7 条染色体,被子植物祖先基因组有 15 条染色体^[32]。和大多数被子植物中检测到的与被子植物家族或更小分支特异性相关的古多倍化事件相比,大多数双子叶植物祖先基因组在进化史中发生了更早的(130~150 MYA)古六倍化事件^[39]。对西瓜(*Citrullus lanatus*)古基因组学研究发现,在其基因组进化中发生了 7 次三倍化,从双子叶植物祖先基因组 7 条染色体经历了 81 次分裂和 91 次融合才形成现存西瓜的 11 条染色体^[53]。凤梨基因组对研究单子叶植物祖先基因组进化具有重要作用,染色体核型进化表明其进化路线开始于单子叶植物祖先基因组的 5 条染色体,经 τ 事件后先出现 10 条染色体后减到 9 条染色体,经 σ 事件后出现 27 条染色体,然后发生 8 次融合和 6 次分裂,最终被整合到现存的 25 条染色体中^[32]。复活草基因组和其他草类基因组共同经历了 ρ WGD 事件^[15],二穗短柄草与小麦、水稻、高粱分化时间分别为 32~39 MYA、40~53 MYA、45~60 MYA,其基因组内部重复分化时间发生在 56~72 MYA,早于草类分化时间^[54]。

整合银杏基因组学^[14]和古生物学^[55]研究表明,银杏可能起源于种子蕨类植物,其基因组经历了 2 次 WGD 事件,先后发生在 515~735 MYA 和 74~147 MYA。被子植物共有银杏较古老的一次 WGD 事件,较近的一次可能是银杏特异性 WGD 事件(图 2)。松柏类植物是裸子植物中庞大的一个类群,具有较大的基因组(20~40 Gb,显著大于银杏基因组 10.61 Gb),约 300 MYA 起源于古老的种子植物^[56]。Li 等^[13]对 22 种裸子植物进行转录组分析,发现裸子植物基因组经历了 3 次 WGD 事件,其中 2 次位于松柏类植物分支(图 2)。在裸子植物进化史中,WGD 事件使松柏类植物与其他裸子植物分化开,松科和柏科均独自经历 1 次 WGD 事件。海岸红杉是柏科中现存的唯一的自然六倍体物种,对其基因组 WGD 事件进行分析,发现两次 WGD 事件,先后发生在 1.5~10 MYA 和 0.4~3 MYA^[52]。

5 展 望

基因组多倍化使植物基因组快速重组,丢失大

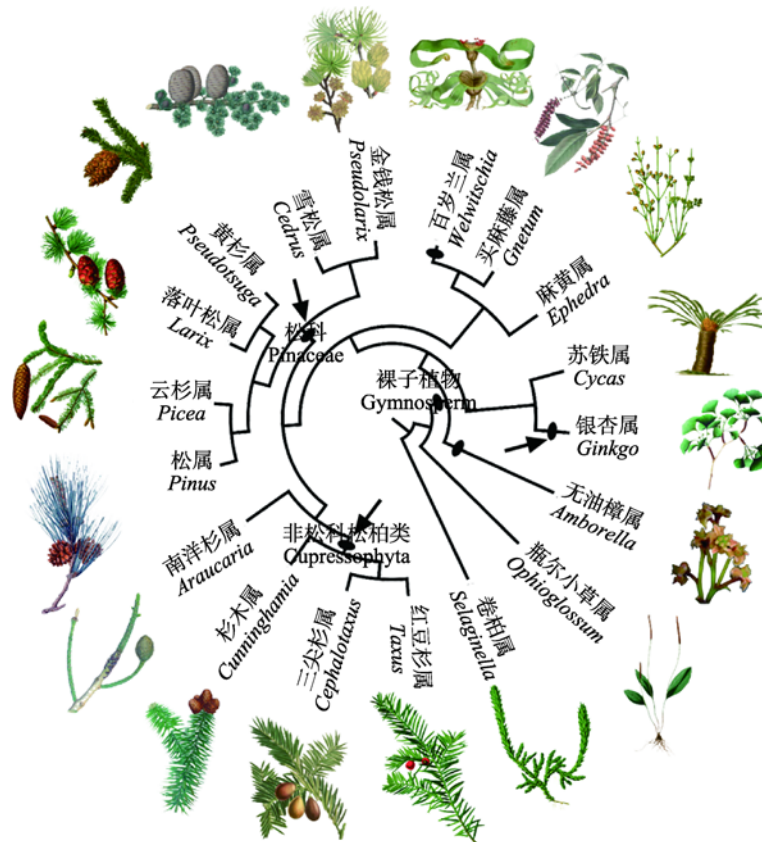


图 2 种子植物与裸子植物 WGD 事件分布简图

Fig. 2 A diagram of WGD events in seed plants and gymnosperm paleohistory

参考文献[13]并修改, 其中银杏 WGD 事件引自文献[14]。黑圈表示 WGD 事件, 箭头所指 3 次 WGD 事件分别由银杏科、松科、非松科松柏类所共享。

量基因, 增加结构变异^[57], 是植物基因组进化的重要动力, 有利于植物适应新的环境^[12,49,58,59]。WGD 事件使被子植物基因组大小增加, 但平均基因组大小与倍性并不相关^[60]。WGD 事件发生后, 常发生大量同源基因对丢失^[61], 使鉴定古老的 WGD 极其困难。通过植物进化重建 WGD 使推断基因组共线数据更加困难, 尤其是使用组装不完整基因组序列检测物种间共线性关系。测序技术在不断革新, 测序长度越来越长, 准确度和完整度逐步提高, 综合运用二代和三代测序平台将为植物古基因组学研究提供更可靠的数据。目前, 已有多个组学大数据管理平台储存和提供可靠的植物全基因组序列资源, 如 JGI(<https://phytozome.jgi.doe.gov/>)和 CoGe(<https://genomevolution.org/coge/>)等在线平台提供更严谨的基因组序列数据, 为古基因组学研究提供可靠资源。1000 种植物转录谱测序项目(1KP 计划)^[62]提供了

1000 种植物转录谱资源, 为比较基因组学提供大量高质量序列资源, 这将有效提升对植物古基因组学的认知。

参考文献(References):

- [1] Abrouk M, Murat F, Pont C, Messing J, Jackson S, Faraut T, Tannier E, Plomion C, Cooke R, Feuillet C, Salse J. Palaeogenomics of plants: synteny-based modelling of extinct ancestors. *Trends Plant Sci*, 2010, 15(9): 479–487. [DOI]
- [2] Salse J, Abrouk M, Murat F, Quraishi UM, Feuillet C. Improved criteria and comparative genomics tool provide new insights into grass paleogenomics. *Brief Bioinform*, 2009, 10(6): 619–630. [DOI]
- [3] Jiao WB, Schneeberger K. The impact of third generation genomic technologies on plant genome assembly. *Curr Opin Plant Biol*, 2017, 36: 64–70. [DOI]

- [4] Xu Q, Chen LL, Ruan XA, Chen DJ, Zhu AD, Chen CL, Bertrand D, Jiao WB, Hao BH, Lyon MP, Chen JJ, Gao S, Xing F, Lan H, Chang JW, Ge XH, Lei Y, Hu Q, Miao Y, Wang L, Xiao SX, Biswas MK, Zeng WF, Guo F, Cao HB, Yang XM, Xu XW, Cheng YJ, Xu J, Liu JH, Luo OJ, Tang ZH, Guo WW, Kuang HH, Zhang HY, Roose ML, Nagarajan N, Deng XX, Ruan YJ. The draft genome of sweet orange (*Citrus sinensis*). *Nat Genet*, 2013, 45: 59–66. [DOI]
- [5] Huang SW, Li RQ, Zhang ZH, Li L, Gu XF, Fan W, Lucas WJ, Wang XW, Xie BY, Ni PX, Ren YY, Zhu HM, Li J, Lin K, Jin WW, Fei ZJ, Li GC, Staub J, Kilian A, van der Vossen EAG, Wu Y, Guo J, He J, Jia ZQ, Ren Y, Tian G, Lu Y, Ruan J, Qian WB, Wang MW, Huang QF, Li B, Xuan ZL, Cao JJ, Wu ZG, Zhang JB, Cai QL, Bai YQ, Zhao BW, Han YH, Li Y, Li XF, Wang SH, Shi QX, Liu SQ, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng ZC, Zhang SP, Wu J, Yang YH, Kang HX, Li M, Liang HQ, Ren XL, Shi ZB, Wen M, Jian M, Yang HL, Zhang GJ, Yang ZT, Chen R, Liu SF, Li JW, Ma LJ, Liu H, Zhou Y, Zhao J, Fang XD, Li GQ, Fang L, Li YR, Liu DY, Zheng HK, Zhang Y, Qin N, Li Z, Yang GH, Yang S, Bolund L, Kristiansen K, Zheng HC, Li SC, Zhang XQ, Yang HM, Wang J, Sun RF, Zhang BX, Jiang SZ, Wang J, Du YC, Li SG. The genome of the cucumber, *Cucumis sativus* L. *Nat Genet*, 2009, 41(12): 1275–1281. [DOI]
- [6] Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, Goodstein DM, Dubchak I, Poliakov A, Mizrahi E, Kullar ARK, Hussey SG, Pinard D, van der Merwe K, Singh P, van Jaarsveld I, Silva-Junior OB, Togawa RC, Pappas MR, Faria DA, Sansaloni CP, Petrolis CD, Yang X, Ranjan P, Tschaplinski TJ, Ye CY, Li T, Sterck L, Vanneste K, Murat F, Soler M, Clemente HS, Saidi N, Cassan-Wang H, Dunand C, Hefer CA, Bornberg-Bauer E, Kersting AR, Vining K, Amarasinghe V, Ranik M, Naithani S, Elser J, Boyd AE, Liston A, Spatafora JW, Dharmawardhana P, Raja R, Sullivan C, Romanell E, Alves-Ferreira M, Kulheim C, Foley W, Carocha V, Paiva J, Kudrna D, Brommonschenkel SH, Pasquali G, Byrne M, Rigault P, Tibbits J, Spokevicius A, Jones RC, Steane DA, Vaillancourt RE, Potts BM, Joubert F, Barry K, Pappas GJ, Strauss SH, Jaiswal P, Grima-Pettenati J, Salse J, Van de Peer Y, Rokhsar DS, Schmutz J. The genome of *Eucalyptus grandis*. *Nature*, 2014, 510(7505): 356–362. [DOI]
- [7] Olsen JL, Rouzé P, Verhelst B, Lin YC, Bayer T, Collen J, Dattolo E, De Paoli E, Dittami S, Maumus F, Michel G, Kersting A, Lauritano C, Lohaus R, Töpel M, Tonon T, Vanneste K, Amirebrahimi M, Brakel J, Boström C, Chovatia M, Grimwood J, Jenkins JW, Jueterbock A, Mraz A, Stam WT, Tice H, Bornberg-Bauer E, Green PJ, Pearson GA, Procaccini G, Duarte CM, Schmutz J, Reusch TBH, Van de Peer Y. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature*, 2016, 530(7590): 331–335. [DOI]
- [8] Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hallman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Kaller M, Luthman J, Lysholm F, Niittyla T, Olson A, Rilakovic N, Ritland C, Rossello JA, Sena J, Svensson T, Talavera-Lopez C, Theissen G, Tuominen H, Vanneste K, Wu ZQ, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Lee Thompson S, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S. The Norway spruce genome sequence and conifer genome evolution. *Nature*, 2013, 497(7451): 579–584. [DOI]
- [9] Peng Z, Lu Y, Li L, Zhao Q, Feng Q, Gao Z, Lu H, Hu T, Yao N, Liu K, Li Y, Fan D, Guo Y, Li W, Lu Y, Weng Q, Zhou C, Zhang L, Huang T, Zhao Y, Zhu C, Liu X, Yang X, Wang T, Miao K, Zhuang C, Cao X, Tang W, Liu G, Liu Y, Chen J, Liu Z, Yuan L, Liu Z, Huang X, Lu T, Fei B, Ning Z, Han B, Jiang Z. The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat Genet*, 2013, 45(4): 456–461. [DOI]
- [10] Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, Zhang L, Niu X, Zhang X, Meng M, Yu J, Liu J, Han Y, Shi W, Zhang D, Cao S, Wei Z, Cui Y, Xia Y, Zeng H, Bao K, Lin L, Min Y, Zhang H, Miao M, Tang X, Zhu Y, Sui Y, Li G, Sun H, Yue J, Sun J, Liu F, Zhou L, Lei L, Zheng X, Liu M, Huang L, Song J, Xu C, Li J, Ye K, Zhong S, Lu BR, He G, Xiao F, Wang HL, Zheng H, Fei Z, Liu Y. Draft genome of the kiwifruit *Actinidia chinensis*. *Nat Commun*, 2013, 4: 2640. [DOI]
- [11] Sankoff D, Zheng C, Zhu Q. The collapse of gene complement following whole genome duplication. *BMC Genomics*, 2010, 11(1): 313. [DOI]
- [12] Hollister JD. Polyploidy: adaptation to the genomic environment. *New Phytol*, 2015, 205(3): 1034–1039. [DOI]
- [13] Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS. Early genome duplications in conifers and other seed plants. *Sci Adv*, 2015, 1(10): e1501084.

[DOI]

- [14] Guan R, Zhao Y, Zhang H, Fan G, Liu X, Zhou W, Shi C, Wang J, Liu W, Liang X, Fu Y, Ma K, Zhao L, Zhang F, Lu Z, Lee SMY, Xu X, Wang J, Yang H, Fu C, Ge S, Chen W. Draft genome of the living fossil *Ginkgo biloba*. *Giga-Science*, 2016, 5(1): 49. [DOI]
- [15] VanBuren R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, Spittle K, Hall R, Gu J, Lyons E, Freeling M, Bartels D, Ten Hallers B, Hastie A, Michael TP, Mockler TC. Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature*, 2015, 527(7579): 508–511. [DOI]
- [16] Michael TP, VanBuren R. Progress, challenges and the future of crop genomes. *Curr Opin Plant Biol*, 2015, 24: 71–81. [DOI]
- [17] Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*, 2016, 13(12): 1050–1054. [DOI]
- [18] Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang ML, Chen J, Biggers E, Zhang J, Huang L, Zhang L, Miao W, Zhang J, Ye Z, Miao C, Lin Z, Wang H, Zhou H, Yim WC, Priest HD, Zheng C, Woodhouse M, Edger PP, Guyot R, Guo HB, Guo H, Zheng G, Singh R, Sharma A, Min X, Zheng Y, Lee H, Gurtowski J, Sedlazeck FJ, Harkess A, McKain MR, Liao Z, Fang J, Liu J, Zhang X, Zhang Q, Hu W, Qin Y, Wang K, Chen LY, Shirley N, Lin YR, Liu LY, Hernandez AG, Wright CL, Bulone V, Tuskan GA, Heath K, Zee F, Moore PH, Sunkar R, Leebens-Mack JH, Mockler T, Bennetzen JL, Freeling M, Sankoff D, Paterson AH, Zhu X, Yang X, Smith JAC, Cushman JC, Paull RE, Yu Q. The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet*, 2015, 47(12): 1435–1442. [DOI]
- [19] Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, Hu Z, Chen S, Pental D, Ju Y, Yao P, Li X, Xie K, Zhang J, Wang J, Liu F, Ma W, Shopan J, Zheng H, Mackenzie SA, Zhang M. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat Genet*, 2016, 48(10): 1225–1232. [DOI]
- [20] Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Brière C, Owens GL, Carrère S, Mayjonade B, Legrand L, Gill N, Kane NC, Bowers JE, Hubner S, Bellec A, Bérard A, Bergès H, Blanchet N, Boniface MC, Brunel D, Catrice O, Chaidir N, Claudel C, Donnadieu C, Faraut T, Fievet G, Helmstetter N, King M, Knapp SJ, Lai Z, Le Paslier MC, Lippi Y, Lorenzon L, Mandel JR, Marage G, Marchand G, Marquand E, Bret-Mestries E, Morien E, Nambeesan S, Nguyen T, Pegot-Espagnet P, Pouilly N, Raftis F, Sallet E, Schiex T, Thomas J, Vandecasteele C, Varès D, Vear F, Vautrin S, Crespi M, Mangin B, Burke JM, Salse J, Muñoz S, Vincourt P, Rieseberg LH, Langlade NB. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, 2017, 546(7656): 148–152. [DOI]
- [21] Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, Bayer M, Ramsay L, Liu H, Haberer G, Zhang XQ, Zhang Q, Barrero RA, Li L, Taudien S, Groth M, Felder M, Hastie A, Šimková H, Staňková H, Vrána J, Chan S, Muñoz-Amatriaín M, Ounit R, Wanamaker S, Bolser D, Colmsee C, Schmutzer T, Aliyeva-Schnorr L, Grasso S, Tanskanen J, Chailyan A, Sampath D, Heavens D, Clissold L, Cao S, Chapman B, Dai F, Han Y, Li H, Li X, Lin C, McCooke JK, Tan C, Wang P, Wang S, Yin S, Zhou G, Poland JA, Bellgard MI, Borisjuk L, Houben A, Doležel J, Ayling S, Lonardi S, Kersey P, Langridge P, Muehlbauer GJ, Clark MD, Caccamo M, Schulman AH, Mayer KFX, Platzer M, Close TJ, Scholz U, Hansson M, Zhang G, Braumann I, Spannagl M, Li C, Waugh R, Stein N. A chromosome conformation capture ordered sequence of the barley genome. *Nature*, 2017, 544(7651): 427–433. [DOI]
- [22] Bertoli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn MC, Shirasawa K, Huang W, Vidigal B, Abernathy B, Chu Y, Niederhuth CE, Umale P, Araujo ACG, Kozik A, Do Kim K, Burow MD, Varshney RK, Wang X, Zhang X, Barkley N, Guimaraes PM, Isobe S, Guo B, Liao B, Stalker HT, Schmitz RJ, Scheffler BE, Leal-Bertoli SCM, Xun X, Jackson SA, Michelsmore R, Ozias-Akins P. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet*, 2016, 48(4): 438–446. [DOI]
- [23] Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, Jung K, Lee GW, Oh SK, Bae C, Kim SB, Lee HY, Kim SY, Kim MS, Kang BC, Jo YD, Yang HB, Jeong HJ, Kang WH, Kwon JK, Shin C, Lim JY, Park JH, Huh JH, Kim JS, Kim BD, Cohen O, Paran I, Suh MC, Lee SB, Kim YK, Shin Y, Noh SJ, Park J,

- Seo YS, Kwon SY, Kim HA, Park JM, Kim HJ, Choi SB, Bosland PW, Reeves G, Jo SH, Lee BW, Cho HT, Choi HS, Lee MS, Yu Y, Do Choi Y, Park BS, van Deynze A, Ashrafi H, Hill T, Kim WT, Pai HS, Ahn HK, Yeam I, Giovannoni JJ, Rose JKC, Sorensen I, Lee SJ, Kim RW, Choi IY, Choi BS, Lim JS, Lee YH, Choi D. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat Genet*, 2014, 46(3): 270–278. [DOI]
- [24] Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol*, 2012, 30(1): 83–89. [DOI]
- [25] Lan T, Renner T, Ibarra-Laclette E, Farr KM, Chang TH, Cervantes-Pérez SA, Zheng C, Sankoff D, Tang H, Purbojati RW, Putra A, Drautz-Moses DI, Schuster SC, Herrera-Estrella L, Albert VA. Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. *Proc Natl Acad Sci USA*, 2017, 114(22): E4435–E4441. [DOI]
- [26] Salojärvi J, Smolander OP, Nieminen K, Rajaraman S, Safronov O, Safdari P, Lamminmaki A, Immanen J, Lan T, Tanskanen J, Rastas P, Amiryousefi A, Jayaprakash B, Kammonen JI, Hagqvist R, Eswaran G, Ahonen VH, Serra JA, Asiegbu FO, de Dios Barajas-Lopez J, Blande D, Blokhina O, Blomster T, Broholm S, Brosche M, Cui F, Dardick C, Ehonen SE, Elomaa P, Escamez S, Fagerstedt KV, Fujii H, Gauthier A, Gollan PJ, Halimaa P, Heino PI, Himanen K, Hollender C, Kangasjarvi S, Kauppinen L, Kelleher CT, Kontunen-Soppela S, Koskinen JP, Kovalchuk A, Karenlampi SO, Karkonen AK, Lim KJ, Leppala J, Macpherson L, Mikola J, Mouhu K, Mahonen AP, Niinemets U, Oksanen E, Overmyer K, Palva ET, Pazouki L, Pennanen V, Puhakainen T, Poczar P, Possen BJHM, Punkkinen M, Rahikainen MM, Rousi M, Ruonala R, van der Schoot C, Shapiguzov A, Sierla M, Sipila TP, Sutela S, Teeri TH, Tervahauta AI, Vaattovaara A, Vahala J, Vetchinnikova L, Welling A, Wrzaczek M, Xu E, Paulin LG, Schulman AH, Lascoux M, Albert VA, Auvinen P, Helariutta Y, Kangasjarvi J. Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nat Genet*, 2017, 49(6): 904–912. [DOI]
- [27] Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Michell CT, Saber N, Kharbatia NM, Rupper RR, Sharp AR, Dally N, Boughton BA, Woo YH, Gao G, Schijlen EGWM, Guo X, Momin AA, Negrão S, Al-Babili S, Gehring C, Roessner U, Jung C, Murphy K, Arold ST, Gojobori T, van der Linden CG, van Loo EN, Jellen EN, Maughan PJ, Tester M. The genome of *Chenopodium quinoa*. *Nature*, 2017, 542(7641): 307–312. [DOI]
- [28] Nowak MD, Russo G, Schlapbach R, Huu CN, Lenhard M, Conti E. The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biol*, 2015, 16(1): 12. [DOI]
- [29] Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, Wang J, Zhou L, Wang D, Gao Y, Wang Y, Shi W, Ji P, Li D, Guan Z, Shao C, Zhuang Z, Gao Z, Qi J, Zhao F. *Saccharina* genomes provide novel insight into kelp biology. *Nat Commun*, 2015, 6: 6986. [DOI]
- [30] Xu H, Song J, Luo H, Zhang Y, Li Q, Zhu Y, Xu J, Li Y, Song C, Wang B, Sun W, Shen G, Zhang X, Qian J, Ji A, Xu Z, Luo X, He L, Li C, Sun C, Yan H, Cui G, Li X, Li Xe, Wei J, Liu J, Wang Y, Hayward A, Nelson D, Ning Z, Peters RJ, Qi X, Chen S. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol Plant*, 2016, 9(6): 949–952. [DOI]
- [31] Salse J. *In silico* archeogenomics unveils modern plant genome organisation, regulation and evolution. *Curr Opin Plant Biol*, 2012, 15(2): 122–130. [DOI]
- [32] Murat F, Armero A, Pont C, Klopp C, Salse J. Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat Genet*, 2017, 49(4): 490–496. [DOI]
- [33] Ohno S. Evolution by gene duplication. Berlin: Springer-Verlag, 1970: 98–106. [DOI]
- [34] Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 1997, 387(6634): 708–713. [DOI]
- [35] Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 2000, 408(6814): 796–815. [DOI]
- [36] Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Huguene P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M,

- Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, Wincker P; French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 2007, 449(7161): 463–467. [DOI]
- [37] Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto SI, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*, 2008, 319(5859): 64–69. [DOI]
- [38] Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, Ashton NW, Axtell MJ, Barker E, Barker MS, Bennetzen JL, Bonawitz ND, Chapple C, Cheng C, Correa LGG, Dacre M, DeBarry J, Dreyer I, Elias M, Engstrom EM, Estelle M, Feng L, Finet C, Floyd SK, Frommer WB, Fujita T, Gramzow L, Gutensohn M, Harholt J, Hattori M, Heyl A, Hirai T, Hiwatashi Y, Ishikawa M, Iwata M, Karol KG, Koehler B, Kolukisaoglu U, Kubo M, Kurata T, Lalonde S, Li K, Li Y, Litt A, Lyons E, Manning G, Maruyama T, Michael TP, Mikami K, Miyazaki S, Morinaga SI, Murata T, Mueller - Roeber B, Nelson DR, Obara M, Oguri Y, Olmstead RG, Onodera N, Petersen BL, Pils B, Prigge M, Rensing SA, Riaño-Pachón DM, Roberts AW, Sato Y, Scheller HV, Schulz B, Schulz C, Shakirov EV, Shibagaki N, Shinohara N, Shippen DE, Sørensen I, Sotooka R, Sugimoto N, Sugita M, Sumikawa N, Tanurdzic M, Theißen G, Ulvskov P, Wakazuki S, Weng JK, Willats WWGT, Wipf D, Wolf PG, Yang L, Zimmer AD, Zhu Q, Mitros T, Hellsten U, Loqué D, Otiillar R, Salamov A, Schmutz J, Shapiro H, Lindquist E, Lucas S, Rokhsar D, Grigoriev IV. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science*, 2011, 332(6032): 960–963. [DOI]
- [39] Albert VA, Barbazuk WB, dePamphilis CW, Der JP, Leebens-Mack J, Ma H, Palmer JD, Rounsley S, Sankoff D, Schuster SC, Soltis DE, Soltis PS, Wessler SR, Wing RA, Albert VA, Ammiraju JSS, Barbazuk WB, Chamala S, Chanderbali AS, dePamphilis CW, Der JP, Determann R, Leebens-Mack J, Ma H, Ralph P, Rounsley S, Schuster SC, Soltis DE, Soltis PS, Talag J, Tomsho L, Walts B, Wanke S, Wing RA, Albert VA, Barbazuk WB, Chamala S, Chanderbali AS, Chang TH, Determann R, Lan T, Soltis DE, Soltis PS, Arikiti S, Axtell MJ, Ayyampalayam S, Barbazuk WB, Burnette JM, Chamala S, De Paoli E, dePamphilis CW, Der JP, Estill JC, Farrell NP, Harkess A, Jiao Y, Leebens-Mack J, Liu K, Mei W, Meyers BC, Shahid S, Wafula E, Walts B, Wessler SR, Zhai J, Zhang X, Albert VA, Carretero-Paulet L, dePamphilis CW, Der JP, Jiao Y, Leebens-Mack J, Lyons E, Sankoff D, Tang H, Wafula E, Zheng C, Albert VA, Altman NS, Barbazuk WB, Carretero-Paulet L, dePamphilis CW, Der JP, Estill JC, Jiao Y, Leebens-Mack J, Liu K, Mei W, Wafula E, Altman NS, Arikiti S, Axtell MJ, Chamala S, Chanderbali AS, Chen F, Chen JQ, Chiang V, De Paoli E, dePamphilis CW, Der JP, Determann R, Fogliani B, Guo C, Harholt J, Harkess A, Job C, Job D, Kim S, Kong H, Leebens-Mack J, Li G, Li L, Liu J, Ma H, Meyers BC, Park J, Qi X, Rajjou L, Burtet-Sarramegna V, Sederoff R, Shahid S, Soltis DE, Soltis PS, Sun YH, Ulvskov P, Villegente M, Xue JY, Yeh TF, Yu X, Zhai J, Acosta JJ, Albert VA, Barbazuk WB, Bruenn RA, Chamala S, de Kochko A, dePamphilis CW, Der JP, Herrera-Estrella LR, Ibarra-Laclette E, Kirst M, Leebens-Mack J, Pissis SP, Poncet V, Schuster SC, Soltis DE, Soltis PS, Tomsho L. The *Amborella* genome and the evolution of flowering plants. *Science*, 2013, 342(6165): 1241089. [DOI]
- [40] McKain MR, Tang H, McNeal JR, Ayyampalayam S, Davis JI, dePamphilis CW, Givnish TJ, Pires JC, Stevenson DW, Leebens-Mack JH. A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biol Evol*, 2016, 8(4): 1150–1164. [DOI]
- [41] Tiley GP, Ané C, Burleigh JG. Evaluating and characterizing ancient whole-genome duplications in plants with gene count data. *Genome Biol Evol*, 2016, 8(4): 1023–1037. [DOI]
- [42] Lyons E, Pedersen B, Kane J, Freeling M. The value of

- nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Trop Plant Biol*, 2008, 1(3): 181–190. [DOI]
- [43] Vanneste K, Van de Peer Y, Maere S. Inference of genome duplications from age distributions revisited. *Mol Biol Evol*, 2013, 30(1): 177–190. [DOI]
- [44] Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, Nelson W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Luo M-C, Thudi M, Gowda CLL, Singh NP, Lichtenzveig J, Gali KK, Rubio J, Nadarajan N, Dolezel J, Bansal KC, Xu X, Edwards D, Zhang G, Kahl G, Gil J, Singh KB, Datta SK, Jackson SA, Wang J, Cook DR. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol*, 2013, 31(3): 240–246. [DOI]
- [45] Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. Genome sequence of the palaeopolyploid soybean. *Nature*, 2010, 463(7278): 178–183. [DOI]
- [46] Cai J, Liu X, Vanneste K, Proost S, Tsai WC, Liu KW, Chen LJ, He Y, Xu Q, Bian C, Zheng Z, Sun F, Liu W, Hsiao YY, Pan ZJ, Hsu CC, Yang YP, Hsu YC, Chuang YC, Dievart A, Dufayard JF, Xu X, Wang JY, Wang J, Xiao XJ, Zhao XM, Du R, Zhang GQ, Wang M, Su YY, Xie GC, Liu GH, Li LQ, Huang LQ, Luo YB, Chen HH, Van de Peer Y, Liu ZJ. The genome sequence of the orchid *Phalaenopsis equestris*. *Nat Genet*, 2015, 47(1): 65–72. [DOI]
- [47] Devos N, Szövényi P, Weston DJ, Rothfels CJ, Johnson MG, Shaw AJ. Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta). *New Phytol*, 2016, 211(1): 300–318. [DOI]
- [48] Wang Y, Ficklin SP, Wang X, Feltus FA, Paterson AH. Large-Scale gene relocations following an ancient genome triplication associated with the diversification of core eudicots. *PLoS One*, 2016, 11(5): e0155637. [DOI]
- [49] del Pozo JC, Ramirez-Parra E. Whole genome duplications in plants: an overview from *Arabidopsis*. *J Exp Bot*, 2015, 66(22): 6991–7003. [DOI]
- [50] Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, Denoeud F, Belcram H, Links MG, Just J, Clarke C, Bender T, Huebert T, Mason AS, Pires JC, Barker G, Moore J, Walley PG, Manoli S, Batley J, Edwards D, Nelson MN, Wang X, Paterson AH, King G, Bancroft I, Chalhoub B, Sharpe AG. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biol*, 2014, 15(6): R77. [DOI]
- [51] Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J-H, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weissshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IAP, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*, 2011, 43(10): 1035–1039. [DOI]
- [52] Scott AD, Stenz NWM, Ingvarsson PK, Baum DA. Whole genome duplication in coast redwood (*Sequoia sempervirens*) and its implications for explaining the rarity of polyploidy in conifers. *New Phytol*, 2016, 211(1): 186–193. [DOI]
- [53] Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z, Min J, Guo X, Murat F, Ham BK, Zhang Z, Gao S, Huang M, Xu Y, Zhong S, Bombarely A, Mueller LA, Zhao H, He H, Zhang Y, Zhang Z, Huang S, Tan T, Pang E, Lin K, Hu Q, Kuang H, Ni P, Wang B, Liu J, Kou Q, Hou W, Zou X, Jiang J, Gong G, Klee K, Schoof H, Huang Y, Hu X, Dong S, Liang D, Wang J, Wu K, Xia Y, Zhao X, Zheng Z, Xing M, Liang X, Huang B, Lv T, Wang J, Yin Y, Yi H, Li R, Wu M, Levi A, Zhang X, Giovannoni JJ, Wang J, Li Y, Fei Z, Xu Y. The

- draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet*, 2013, 45(1): 51–58. [DOI]
- [54] Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smoth M, Lail K, Tice H, Grimwood J, McKenzie N, Huo N, Gu YQ, Lazo GR, Anderson OD, You FM, Luo MC, Dvorak J, Wright J, Febrer M, Idziak D, Hasterok R, Lindquist E, Wang M, Fox SE, Priest HD, Filichkin SA, Givan SA, Bryant DW, Chang JH, Wu H, Wu W, Hsia AP, Schnable PS, Kalyanaraman A, Baarbazuk B, Michael TP, Hazen SP, Bragg JN, Laudencia-Chingcuanco D, Weng Y, Haberer G, Spannagl M, Mayer K, Rattei T, Mitros T, Lee SJ, Rose JKC, Mueller LA, York TL, Wicker T, Buchmann JP, Tanskanen J, Schulman AH, Gundlach H, de Oliveira AC, Maia LdC, Belknap W, Jiang N, Lai J, Zhu L, Ma J, Sun C, Pritham E, Salse J, Murat F, Abrouk M, Bruggmann R, Messing J, Fahlgren N, Sullivan CM, Carrington JC, Chapman EJ, May GD, Zhai J, Ganssmann M, Gurazada SGR, German M, Meyers BC, Green PJ, Tyler L, Wu J, Thomson J, Chen S, Scheller HV, Harholt J, Ulvskov P, Kimbrel JA, Bartley LE, Cao P, Jung KH, Sharma MK, Vega-Sanchez M, Ronald P, Dardick CD, De Bodt S, Verelst W, Inze D, Heese M, Schnittger A, Yang X, Kalluri UC, Tuskan GA, Hua Z, Vierstra RD, Cui Y, Ouyang S, Sun Q, Liu Z, Yilmaz A, Grotewold E, Sibout R, Hematy K, Mouille G, Hoefte H, Pelloux J, O'Connor D, Schnable J, Rowe S, Harmon F, Cass CL, Sedbrook JC, Byrne ME, Walsh S, Higgins J, Li P, Brutnell T, Unver T, Budak H, Belcram H, Charles M, Chalhoub B, Baxter I. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, 2010, 463(7282): 763–768. [DOI]
- [55] Herrera F, Shi G, Ichinnorov N, Takahashi M, Bugdaeva EV, Herendeen PS, Crane PR. The presumed ginkgophyte *Ulmalepis* has seed-bearing structures resembling those of Peltaspermales and Umkomasiales. *Proc Natl Acad Sci USA*, 2017, 114(12): E2385–E2391. [DOI]
- [56] Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, Martínez-García PJ, Vasquez-Gross HA, Lin BY, Zieve JJ, Dougherty WM, Fuentes-Soriano S, Wu LS, Gilbert D, Marçais G, Roberts M, Holt C, Yandell M, Davis JM, Smith KE, Dean JF, Lorenz WW, Whetten RW, Sederoff R, Wheeler N, McGuire PE, Main D, Loopstra CA, Mockaitis K, deJong PJ, Yorke JA, Salzberg SL, Langley CH. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol*, 2014, 15(3): R59. [DOI]
- [57] Chaney L, Sharp AR, Evans CR, Udall JA. Genome mapping in plant comparative genomics. *Trends Plant Sci*, 2016, 21(9): 770–780. [DOI]
- [58] Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Plant Biol*, 2016, 30: 159–165. [DOI]
- [59] Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. *Plant Physiol*, 2016, 171(4): 2294–2316. [DOI]
- [60] Zenil-Ferguson R, Ponciano JM, Burleigh JG. Evaluating the role of genome downsizing and size thresholds from genome size distributions in angiosperms. *Am J Bot*, 2016, 103(7): 1175–1186. [DOI]
- [61] Sankoff D, Zheng CF, Zhu QA. The collapse of gene complement following whole genome duplication. *BMC Genomics*, 2010, 11: 313. [DOI]
- [62] Matasci N, Hung LH, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, Burleigh JG, Gitzendanner MA, Wafula E, Der JP, dePamphilis CW, Roure B, Philippe H, Ruhfel BR, Miles NW, Graham SW, Mathews S, Surek B, Melkonian M, Soltis DE, Soltis PS, Rothfels C, Pokorny L, Shaw JA, DeGironimo L, Stevenson DW, Villarreal JC, Chen T, Kutchan TM, Rolf M, Baucom RS, Deyholos MK, Samudrala R, Tian Z, Wu X, Sun X, Zhang Y, Wang J, Leebens-Mack J, Wong GKS. Data access for the 1, 000 Plants (1KP) project. *GigaScience*, 2014, 3: 17. [DOI]

(责任编辑: 储成才)