

DOI: 10.13376/j.cbls/2022034

文章编号: 1004-0374(2022)03-0285-09

植物着丝粒组成和功能的研究进展

徐美玲, 孙健英*, 李宗芸*

(江苏师范大学生命科学学院, 徐州 221116)

摘要: 着丝粒是由 DNA 与蛋白质组成的复合物, 是染色体上的重要功能位点, 在细胞有丝分裂和减数分裂过程中参与动粒的组装, 控制染色单体的配对与分离。真核生物着丝粒 DNA 含有大量重复序列且序列进化迅速, 物种间相似性较低, 因而着丝粒区域 DNA 的组装成为全基因组测序中最困难的任务之一。植物着丝粒特异性组蛋白 CENH3 可以定义活性着丝粒, 其作为表观遗传标记以建立和维持着丝粒的功能。随着染色质免疫共沉淀、测序技术的发展并结合分子生物学、细胞遗传学、生物信息学的分析手段, 我们可以进一步对着丝粒序列进行分析及定位, 同时为其功能和进化等研究提供一定的依据。该文将重点介绍植物着丝粒的组成、功能、表观遗传现象、研究技术和测序方法等。

关键词: 着丝粒; 染色体; 动粒; 组蛋白; 着丝粒序列

中图分类号: Q946.5; Q591.4

文献标志码: A

Research progress on the composition and function of plant centromeres

XU Mei-Ling, SUN Jian-Ying*, LI Zong-Yun*

(School of Life Sciences, Jiangsu Normal University, Xuzhou 221116, China)

Abstract: The centromere consisting of DNA and protein is an essential functional chromosome domain that participates in kinetochore formation and mediates faithful chromatids pairing and separation during mitosis and meiosis. Universally, the centromere DNA of eukaryotes is enriched with repetitive sequences that evolve rapidly and have few similarities among species. Therefore, the assembly of centromere DNA becomes one of the most difficult tasks in genome-wide sequencing. Plant centromere-specific histone H3 variant (CENH3) can define active centromere, which is used as an epigenetic marker to establish and maintain the function of centromere. As chromatin immunoprecipitation and sequencing technology develop rapidly with the analysis methods of molecular biology, cytogenetics and bioinformatics, we can further analyze and locate centromere sequences and investigate its function and evolution. This review will focus on the research progress of the composition, function, epigenetic phenomenon, and sequencing method of plant centromeres.

Key words: centromere; chromosome; kinetochore; histone; centromeric sequence

1 着丝粒的结构与保守功能

19 世纪 80 年代初, Flemming 等在实验过程中观察到染色体的特定部位有明显的缢痕——主缢痕

(primary constriction), 并首次描述着丝粒 (centromere)^[1-2]。

着丝粒结构复杂, 由外到内包含动粒结构域、中央结构域和配对结构域 3 个部分^[3]。着丝粒的两侧各

收稿日期: 2021-09-21; 修回日期: 2022-01-15

基金项目: 江苏省高等学校自然科学研究面上项目(20KJB180009); 国家自然科学基金项目(32000376, 31771367); 财政部和农业农村部国家现代农业产业技术体系(甘薯)项目(CARS-10-B02); 江苏师范大学研究生科研与实践创新计划项目(2021XKT0782)

*通信作者: 孙健英, Tel: 0516-83656616, E-mail: jianyingsun@jsnu.edu.cn; 李宗芸, Tel: 0516-83656616, E-mail: zongyunli@jsnu.edu.cn

附着一个两层蛋白质的盘状结构,称为动粒(kinetochore),动粒外层与纺锤丝微管连接,动粒内层与着丝粒核心区染色质紧密结合^[4]。着丝粒与动粒在结构和功能上关联密切,在染色体上能够与动粒发生相互作用的片段称为功能着丝粒^[5-6]。

根据着丝粒在物种间的大小和分布不同,可以将其分为3种类型:点着丝粒、区域着丝粒、全着丝粒^[7]。例如,芽殖酵母的点着丝粒的大小为125 bp。而在高等植物中,着丝粒的类型多数为区域着丝粒,大小通常达到 kb 至 Mb^[8]。一般情况下,高等植物中一条染色体上只存在一个着丝粒,根据着丝粒在染色体上的位置不同将其分为中间着丝粒(metacentric)、近中间着丝粒(submetacentric)、端着丝粒(telocentric)、近端着丝粒(acrocentric)。

着丝粒是真核生物在进行有丝分裂或减数分裂时遗传物质向下一代传递的重要结构区域,主要起到两个关键作用:一是调节姐妹染色单体的黏合和分离;二是参与动粒组装,作为纺锤体微管附着位点,可以监测微管连接的完整性^[9-10]。着丝粒能保持染色体的稳定,维系真核生物的生长与遗传,一旦着丝粒区域发生部分丢失,则可能导致细胞中染色体的结构和数目发生改变^[10-11]。

2 植物着丝粒的基本组成

植物着丝粒的基本组成主要包括着丝粒 DNA 序列和着丝粒蛋白。

2.1 着丝粒DNA序列

2.1.1 着丝粒串联重复序列

着丝粒串联重复序列(tandem repeat, TR)常含有多个拷贝形成卫星序列^[11],是着丝粒 DNA 的主要成分,其通过不等交换、复制跳跃、转座和剪切等一系列突变过程进行快速进化^[12]。经研究发现,大部分卫星重复序列在不同物种之间相似性极低,甚至在近缘物种间组成差异也较大^[13-14]。

卫星序列重复单元的长度一般介于 150 bp~200 bp 之间,约为一个核小体 DNA 的长度,并且倾向于富含 AT。例如拟南芥着丝粒中的卫星重复序列 pAL1 长度为 178 bp^[15],玉米的 CentC 为 156 bp^[16],水稻的 CentO 为 155 bp^[17]。但在大豆中卫星序列长度仅有 92 bp^[18],而马铃薯中长度达到 1 kb~5 kb^[19]。即使是在同一个物种中,卫星 DNA 序列长度也会有不同^[20]。此外,还有一些植物着丝粒卫星 DNA 大小也被报道过,部分如下表(表 1)。

2.1.2 着丝粒反转录转座子

着丝粒上特异的反转录转座子(centromeric retrotransposon, CR)是由 RNA 介导的转座元件,它通过转录合成 RNA 后,以 mRNA 为模板,在反转录酶作用下逆转录成多拷贝 DNA 序列散布在着丝粒串联重复序列之中,这种有规律的穿插分布可能形成了一个较为稳定的 DNA 构象,有利于 CENH3 核小体组装^[35-36]。

与着丝粒串联重复序列相比,反转录转座子在植物着丝粒中含量较少,进化较慢。在高等植物中,着丝粒特异的反转录转座子序列主要是长末端家族(LTR)中的 *Ty3/gypsy* 和 *Ty1/copia* 类型,它们都包含长末端重复序列、pol 开放阅读框、内部 gag 等结构(图 1)^[37]。两者的区别是编码蛋白质的基因顺序不同,*Ty3/gypsy* 家族高度特异地分布在着丝粒区域,而 *Ty1/copia* 家族则不具有着丝粒特异性,在整条染色体上都有分布^[38]。大部分植物的着丝粒反转录转座子属于 *Ty3/gypsy* 类型,而在小麦、黑芥和古莲着丝粒中主要是 *Ty1/copia* 家族类反转录转座子^[39-40]。

表1 已报道部分植物的着丝粒卫星 DNA 序列重复单元大小

物种	卫星DNA序列重复单元长度(bp)	参考文献
<i>Arabidopsis arenosa</i>	166~179	[21]
<i>Arabidopsis thaliana</i>	178	[15]
<i>Aegilops speltoides</i>	250	[22]
<i>Arabidopsis gemmifera</i>	180	[23]
<i>Arachis batizocoi</i>	297	[24]
<i>Astragalus sinicus</i>	20	[25]
<i>Beta corolliflora</i>	162	[26]
<i>Brachypodium distachyon</i>	156	[27]
<i>Brassica rapa</i>	176	[17]
<i>Brassica napus</i>	176	[28]
<i>Brassica oleracea</i>	171	[29]
<i>Brassica campestris</i>	175	[29]
<i>Coix aquatica</i>	153	[30]
<i>Coix lacryma-jobi</i>	153	[30]
<i>Medicago truncatula</i>	166	[31]
<i>Oryza sativa</i>	155	[17]
<i>Oryza brachyantha</i>	154	[32]
<i>Saccharum officinarum</i>	140	[33]
<i>Solanum tuberosum</i>	979~5 400	[19]
<i>Triticum aestivum</i>	570	[34]
<i>Zea mays</i>	156	[16]

最早在禾本科植物中发现的 CR 家族 (短柄草属中的序列 CCS1 和高粱属中的序列 pSau3A9), 是已知的最保守的转座元件之一^[40], 广泛存在于禾谷类植物的着丝粒中。经深入研究发现, 这些序列都属于 *Ty3/gypsy* 类型的反转录转座子, 它们的序列同源性高达 80%^[41], 例如水稻着丝粒中的 CRR、玉米的 CRM、甘蔗的 CRS、大麦的 Cereba、黑麦的 Bilby、野生一粒小麦的 CRW 等。表 2 列举了目前已知的部分植物着丝粒反转录转座子。

2.1.3 功能基因

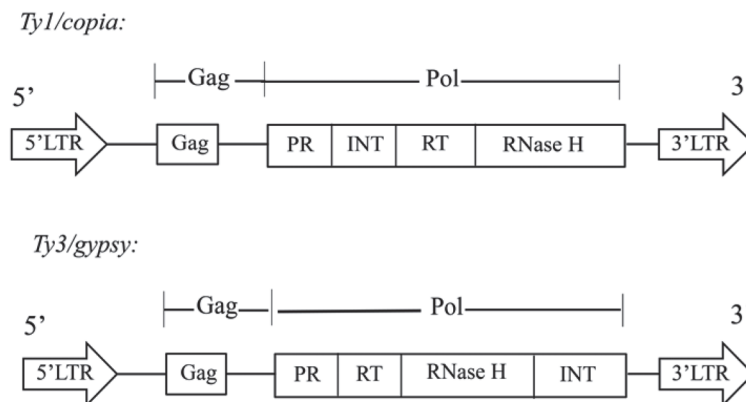
由于着丝粒区域含有大量的重复序列, 且组蛋白 H3 第 9 位赖氨酸常常发生双甲基化修饰而形成异染色质, 因而遗传重组受到一定的抑制^[51], 但着丝粒区域还存在部分低拷贝 DNA 序列以及基因, 其中极少数基因具有转录活性^[52]。例如, 在雷蒙德氏棉^[44]、马铃薯^[19]、中国古莲^[40]、拟南芥 4 号和 8 号染色体^[53]、水稻 8 号染色体^[54] 的着丝粒区域都已经鉴定出有活性基因存在, 个别基因具有性别决定

以及广谱抗病性, 表明着丝粒除了在细胞分裂中发挥作用外, 还可能具备其他重要的生物学作用^[55-56]。

2.2 植物着丝粒蛋白

核小体是构成真核生物染色体的基本结构单位, 由 DNA 和组蛋白 (H2A、H2B、H3、H4) 组成。依据着丝粒蛋白在细胞分裂时存在时间的长短, 可以将它们分为瞬时蛋白质和基础蛋白质^[57]。瞬时蛋白质一般位于动粒蛋白复合体的外层, 染色体分离时开始组装, 分离结束后便解体; 而基础蛋白质一般分布于内层, 始终与着丝粒 DNA 结合^[57]。

在植物着丝粒蛋白的研究中, 目前发现 CENH3 蛋白是着丝粒蛋白复合体最为关键的蛋白 (图 2)。CENH3 是组蛋白 H3 的变体, 可以与 H4、H2A、H2B 组成 CENH3 核小体, 该核小体可以触发着丝粒结合网络 (constitutive centromere associated network, CCAN) 的组装。CCAN 是动粒复合体的一个亚复合体, 它与着丝粒结合, 并招募另一个亚复合体 KMN (KNL1、Mis12 和 Ndc80 复合体) 网络, 与纺



注: LTR: 长末端重复序列; Gag: 编码外壳蛋白基因; PR: 蛋白酶; INT: 整合酶; RT: 反转录酶

图1 *Ty1/copia*与*Ty3/gypsy*反转录转座子的结构示意图^[40]

表2 部分植物的着丝粒反转录转座子

物种	反转录转座子	参考文献
<i>A.thaliana</i>	Athila	[42]
<i>Beta vulgaris</i>	pBv26	[43]
<i>Beta procumbens</i>	pBp10	[26, 43]
<i>Gossypium hirsutum</i>	CRG	[44]
<i>Hordeum vulgares</i>	Cereba	[45]
<i>O.sativa</i>	CRR	[46]
<i>Secale cereale</i>	Bilby	[47]
<i>Sorghum bicolor</i>	pSau3A9	[48]
<i>S.officinamm</i>	CRS	[35]
<i>Zingeria biebersteiniana</i>	Zb47A	[49]
<i>Z.mays</i>	CRM	[50]

锤体微管结合^[58]。在活性着丝粒区域中, CENH3 核小体与 H3 核小体穿插间隔排布, 但 H3 核小体数量居多^[57]。从结构上比较, CENH3 和 H3 蛋白的 C 端都含有保守的组蛋白折叠域 (HFD), HFD 区域包含着丝粒定位信息 CATD, CENH3 的特异性则来自于它序列和长度高度变异的 N 末端^[58]。由于 CENH3 核小体 DNA 的结构更宽松, 所以较 H3 更容易募集到着丝粒区域。CENH3 蛋白对于着丝粒功能的行使是必需的, 因而 CENH3 蛋白可作为活性着丝粒的标记。目前 CENH3 主要具有 3 个方面的应用^[59-61]: (1) 鉴定着丝粒 DNA 序列; (2) 确定功能着丝粒边界和大小; (3) 介导亲本一方基

因组的消除来获得单倍体。

3 着丝粒DNA的起源与进化

真核生物着丝粒的功能十分保守，但 DNA 序列进化快速，常表现出种属特异性^[37]。研究着丝粒 DNA 序列可以为阐明植物着丝粒的组成、功能和进化机制提供线索^[62]。对于解释着丝粒序列在不同物种间快速趋异演化的现象，主要有以下几种假说：第一种假说是减数分裂驱动假说 (meiotic drive hypothesis)，在不对称雌性减数分裂中，着丝粒序列的差异暴露在选择压下，受分子驱动使得重复序列进行扩增，进而 DNA 序列快速进化，甚至产生生殖隔离，最终导致物种的形成^[63-65]。第二种假说是库假说 (library hypothesis)，着丝粒自身本来就存在许多不同丰度的重复序列，在突变等作用下导致新的重复序列的产生，借助遗传漂变和不均衡重组实现着丝粒序列的快速演化^[66]。第三种假说认为着丝粒 DNA 序列的进化和维持与着丝粒所受到的功能压力相关^[13]。第四种假说则认为着丝粒重复序列没有经过重复单元的逐步积累，而是在演化过程中突然出现的^[20]。

除了以上 4 种假说外，另外有研究发现一些物种的着丝粒区域卫星 DNA 和反转录转座子的序列存在相似性，例如玉米^[67]等，推测一部分卫星序列可能由转座子衍生而来，它们在着丝粒区域呈现适应性进化^[37]。此外，在拟南芥^[68]、水稻^[69]、马铃薯^[70]、中国古莲^[40]等中已多次报道着丝粒 DNA 序列和端粒 DNA 序列具有相似性，推测有些着丝粒 DNA 序列可能来源于端粒重复序列^[71]。

4 着丝粒的表观遗传现象

在某些情况下，一条染色体上可能产生两个或多个着丝粒，但进化过程中趋向于只有一个功能着丝粒，其他着丝粒会失活从而保持染色体稳定^[72]。新着丝粒能够在染色体的新位点出现，并在分裂时稳定存在且具有正常的功能，这表明高等生物着的丝粒具有不受 DNA 序列决定的表观遗传特点。

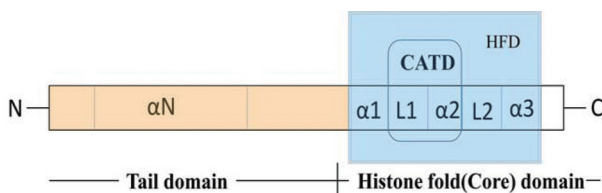


图2 着丝粒组蛋白H3蛋白(CENH3)的结构特征^[61]

表观遗传学主要有 DNA 甲基化、基因印迹和组蛋白修饰等机制。DNA 甲基化主要参与基因表达调控，基因印迹是指不遵从孟德尔定律而依靠单亲传递的某些遗传学性状，而组蛋白修饰是指组蛋白在相关酶作用下发生甲基化、乙酰化、磷酸化、泛素化、腺苷酸化等修饰的过程。在着丝粒 DNA 的表观遗传修饰方面，拟南芥和玉米着丝粒 CENH3 染色质中的卫星序列低甲基化，而周着丝粒区域高甲基化^[73]。在玉米内源 B 染色体中，低甲基化着丝粒 DNA 功能正常，而着丝粒 DNA 高甲基化导致着丝粒失活^[74]。组蛋白修饰对于着丝粒的结构和功能尤为重要，主要发生在组蛋白 N 末端^[75]。不同的组蛋白修饰能够决定染色质具有不同的状态。例如，组蛋白 H3 在 Lys4 位点的甲基化 (H3K4me) 是常染色质状态的标志，基因可以正常表达。而组蛋白 H3 在 Lys9 位点的 H3K9me2 (组蛋白 H3 第 9 位赖氨酸二甲基化) 和 H3K9me3 (组蛋白 H3 第 9 位赖氨酸三甲基化) 这两种修饰是异染色质状态的标志，抑制基因转录^[76]。对部分植物的研究证实，不活跃的着丝粒缺乏 CENH3 和常染色质的组蛋白修饰 (H3K9ac、H3K14ac、H3K4me2、H3K36me2) 以及磷酸化 H3(PH3S10) 修饰等^[77]。相反，不活跃的区域着丝粒富含异染色质相关的表观遗传标记 (H3K9me2/3 和 H3K27me2/3) 和 DNA 甲基化^[78]。此外，异染色质成分的突变并不能消除表观遗传的着丝粒失活，但会导致着丝粒重新激活的比率更高^[79]。

5 着丝粒的转录

过去几十年的研究对着丝粒的经典描述“着丝粒是稳定的、不变的、转录沉默的染色体成分”提出了挑战。基于对包括酵母、真菌、植物和动物在内的广泛真核物种的研究，发现着丝粒区域以细胞周期特异性的方式协调蛋白质复合物组装、染色质组装和转录活动。最近的研究表明，特定类型核酸的转录是决定着丝粒完整性和功能的关键过程^[80]。

多数着丝粒 RNA 是由 RNA Pol II 在严格控制的、相对较低的水平上转录的。对芽殖酵母和人类人工染色体的研究都表明，只有适度的转录水平才能与正确的着丝粒功能相容，这表明这一过程存在严格的控制^[81-82]。因而尽管着丝粒内可以容忍一定水平的转录，但着丝粒的可塑性是有限的，强转录和超乙酰化的组合与活跃的着丝粒功能是不相容的。

着丝粒转录在小鼠卫星 DNA 中首次报道^[83]。有趣的是，着丝粒核心和侧翼周围区域，两个主要

的着丝粒结构域都有转录现象存在。植物中的研究揭示着丝粒上的转录对于 CENH3 的组装以及着丝粒的形成和维持是必不可少的^[84]。转录活性可能起到招募染色质修饰物来调节着丝粒染色质环境的作用, 或者着丝粒 RNA 可能起到促进动粒作为结构或调节成分的作用^[85]。在着丝粒核心及周围异染色质的转录、RNAi 途径和特异性的组蛋白修饰可能协同作用, 促进姐妹染色单体的凝聚、CENH3 的装载或着丝粒的建立^[86]。对玉米的研究表明, 着丝粒转录的启动子可能存在于着丝粒的反转录转座子元件中^[87]。然而, 这是否适用于所有着丝粒转录, 或者转录因子是否与着丝粒内的逆转录元件结合尚不清楚。使用缺乏 α 卫星序列的新着丝粒和含有两个 α 卫星序列结构的双着丝粒染色体的工作表明, RNA pol II 在有丝分裂过程中定位于动粒形成的活性位点, 表明有丝分裂着丝粒转录与活跃的动粒连锁^[88]。

综上所述, 着丝粒转录已进化为着丝粒染色体结构域的基本特征, 着丝粒的转录行为比如通过 RNA Pol II 的作用而触发染色质重塑事件或产生的转录产物 (lncRNA、cenRNA) 可能在着丝粒功能中发挥一定的作用。

6 植物着丝粒的研究方法

6.1 着丝粒蛋白 CENH3 的研究

着丝粒功能十分保守, 但着丝粒蛋白 CENH3 一级结构在物种间保守性较差, 导致 CENH3 抗体很难在物种间通用^[89]。Foltz 等^[90]提取细胞染色质后, 利用串联亲和纯化技术得到 CENH3 核小体, 再使用质谱技术获得了 7 个 CENP 组分。组蛋白 H3 的 CENH3 突变体是研究 CENH3 蛋白功能的重要工具, 在植物中对 CENH3 进行敲除和过表达等操作, 可以探究 CENH3 对细胞分裂和植物生长发育的影响^[91-92]。例如, 在拟南芥中, RNAi 介导的 CENH3 敲除损害了着丝粒功能, 导致有丝分裂以及减数分裂细胞的数量异常和减少, 并且植株生长矮小^[93]。Kelliher 等^[94]使用 RNAi 方法或通过插入转座子来沉默 CENH3 成功产生玉米的单倍体。此外, 还利用改良的 CENH3 对这些玉米品系进行了遗传转化。用转基因植株的花粉为野生型雌花序授粉, 可获得母本单倍体。但 RNAi 基因敲除 CENH3 使得后代单倍体频率很低。CENH3 介导的一个单倍体诱导系就足以获得一个物种的单倍体, 并且产生的单倍体是非转基因的, 可以直接用于植物育种, 而不会引起生物安全的担忧^[95]。

6.2 着丝粒 DNA 序列的研究

植物着丝粒区域含有较多的重复序列, 因而鉴别和组装 DNA 序列面临一定的困难。传统的分析方法主要有超高速离心、染色体显微切割、筛选着丝粒特异 BAC 克隆、限制性内切酶酶切等, 但是由于这些实验操作较复杂, 序列拼接困难, 所以使用频率逐渐降低^[96]。

鉴于植物 CENH3 蛋白具有与着丝粒 DNA 序列直接结合的特点, 近些年来出现的染色质免疫共沉淀的 DNA-蛋白质互作方法结合高通量测序技术 (ChIP-seq) 可以有效获得着丝粒区域的序列信息^[97]。实验技术流程: 首先通过生物信息学方法找到能够区别于普通 H3 蛋白的着丝粒特异 CENH3 蛋白的序列, 然后制备能够识别着丝粒蛋白 CENH3 的特异性抗体, 通过免疫染色实验验证抗体的有效性, 即能否在染色体的主缢痕处产生单一明亮的信号。其次, 为了获得植物功能着丝粒的 DNA 片段, 利用 ChIP 实验将 CENH3 蛋白和对应的 DNA 沉淀下来, 接着解交联释放着丝粒 DNA 片段, 构建 ChIP-DNA 文库, 然后对这些片段进行测序^[98]。基于 Novák 等^[99]开发的 RepeatExplorer 平台, 可以通过图形聚类算法分析全基因组测序数据来识别高等植物基因组中的大多数高拷贝和中拷贝重复。Neumann 等^[100]使用 RepeatExplorer 鉴定的重复序列作为参考, 利用 ChIP-Seq Mapper 评估来自 ChIP 实验的测序数据中重复序列的富集, 在聚类群中选出富集度 (ChIP/Input) 比率最高和重复序列总长最长的重叠群作为着丝粒候选重复序列, 结合荧光原位杂交技术 (fluorescence *in situ* hybridization, FISH) 识别并定位着丝粒区域的 DNA 重复序列。利用 ChIP-Seq 研究着丝粒 DNA 具有分辨率高、覆盖率高等优点, 但 ChIP 实验设计是否合理、抗体制备的效果、染色质片段程度以及测序深度可能会对实验结果产生一定影响^[101]。此外, 夏强明等^[102]利用半四分体分析法 (half-tetrad analysis, HTA) 和 54 株 2n 雌配子有性杂种四倍体后代定位了柑橘着丝粒的物理位置并估计着丝粒候选区域的大小。同时, 利用 ChIP-seq 技术获得与柑橘 CENH3 结合的着丝粒 DNA 序列, 筛选的着丝粒重复序列经 FISH 验证在柑橘染色体着丝粒处有荧光信号, 表明结合这 2 种方法可以有效定位着丝粒区域。

大多数真核生物的基因组组装都处于碎片化状态, 组装的并不完全, 特别是着丝粒等含有大量重复序列的区域。然而, 随着长读测序的最新进展,

包括牛津纳米孔技术 (Oxford nanopore technology, ONT) 和 PacBio 单分子技术, 已经给该领域带来了革命性的变化, 使得着丝粒序列的组装也成为可能。由于它们增加了长度和准确度 (长度为 10 kb 到 >100 kb, 平均准确率为 90%~99%), 长读能够跨越和组装重复序列, 可以实现更完整和连续的基因组组装^[103]。

Song 等^[103] 结合高覆盖率和准确度的长读序列数据和多种组装策略, 填补基因组的组装差距, 获得籼稻品种 ZS97 和 MH63 两个无间隙基因组组合, 这是公开获得的第一个无间隙植物基因组组合。无间隙基因组序列的获得为全球分析和比较两个水稻品种所有染色体的着丝粒提供机会。该研究在水稻着丝粒区域至少鉴定了 395 个非转座元件 (non-TE) 基因, 其中 41% 被发现是活跃转录的。这些测序和分析更新了人们对水稻基因组整体结构和功能的看法。Naish 等^[104] 使用超长读测序获得了 Col-CEN 拟南芥基因组组件, 解析了拟南芥所有着丝粒, 同时推测拟南芥着丝粒是通过卫星序列同质化和反转录转座子插入驱动的多样化进化的。

此外, 基于 Vondrak 等^[105] 探索的一种利用牛津纳米孔技术推断卫星重复序列阵列的分布、查找重复序列间的关联以及推测主要的序列周期的方法, Yang 等^[106] 利用 Hi-C 热图数据中着丝粒的绝缘特性, 成功地开发了一种基于纳米孔长读序列组装和 Hi-C 3D 基因组数据的着丝粒表征新方法, 精确定位了棉花着丝粒。该方法首先将 Hi-C 表征数据与参考基因组进行映射, 获得有效读长进而生成 Hi-C 热图, 使用该热图查找染色体臂内相互作用存在障碍的区域; 接着基于系统发育关系, 使用已知的棉花着丝粒 LTR 与参考基因组比对以验证 Hi-C 着丝粒; 最终得到的着丝粒序列特征, 包括序列组成、LTR 插入时间、LTR 插入模式等可以用于研究着丝粒进化关系。Yang 等^[106] 通过该方法还分析了两个棉花物种的基因组进化, 揭示了棉花 D 基因组虽然在基因含量上是保守的, 但在基因结构、3D 染色质结构、远程调控和胁迫性状上是多样化的。该研究加深了对棉花物种着丝粒进化差异的理解并揭示了其三维基因组结构, 可以更好地指导基础研究和应用来改良作物。

7 未来与展望

与其他真核生物一样, 植物着丝粒是高度复杂和动态的结构^[107]。对其操作和遗传的控制涉及表观遗传现象, 这增加了着丝粒研究的兴趣和难度^[108-109]。

近几十年来, 新一代测序技术、染色质免疫沉淀技术、蛋白质免疫定位技术和荧光原位杂交技术的发展使得植物着丝粒生物学研究取得了显著进展, 包括着丝粒结构、DNA 序列、功能、起源、进化、基因组学、表观遗传学等方面^[110-113]。未来使用植物人工染色体将有助于研究着丝粒长度、卫星序列组成和着丝粒结合蛋白在减数分裂和有丝分裂过程中的作用。然而, 仍有很多关键问题需要去探讨和研究。因此, 继续增加研究物种的数量是非常重要的, 特别是有测序基因组的物种, 这样有助于我们更快地对着丝粒这种维持生命至关重要的有趣结构构建一个更具说明性的图谱。

[参 考 文 献]

- [1] Flemming W. Beitrage zur Kenntniss der Zelle und ihrer Lebenserscheinungen. Archiv für Mikroskopische Anatomie, 1879, 16: 302-436
- [2] Talbert PB, Henikoff S. What makes a centromere? Exp Cell Res, 2020, 389: 111895
- [3] Price CM. Centromeres and telomeres. Curr Opin Cell Biol, 1992, 4: 379-84
- [4] Navarro AP, Cheeseman IM. Kinetochores throughout the cell cycle. Semin Cell Dev Biol, 2021, 117: 62-74
- [5] Cleveland DW, Mao YH, Sullivan KF. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell, 2003, 112: 407-21
- [6] Prosée RF, Wenda JM, Steiner FA. Adaptations for centromere function in meiosis. Essays Biochem, 2020, 64: 193-203
- [7] Ekwall K. Epigenetic control of centromere behavior. Annu Rev Genet, 2007, 41: 63-81
- [8] Sun X, Wahlstrom J, Karpen G. Molecular structure of a functional *Drosophila* centromere. Cell, 1997, 91: 1007-19
- [9] Biggins S, Murray AW. Sister chromatid cohesion in mitosis. Curr Opin Cell Biol, 1998, 10: 769-75
- [10] Wu Y, Kikuchi S, Yan H, et al. Euchromatic subdomains in rice centromeres are associated with genes and transcription. Plant Cell, 2011, 23: 4054-64
- [11] Leo L, Marchetti M, Giunta S, et al. Epigenetics as an evolutionary tool for centromere flexibility. Genes (Basel), 2020, 11: 809
- [12] 陈晔, 刘玉玲, 韦洋洋, 等. 瑟伯氏棉着丝粒重复序列鉴定. 河南师范大学学报(自然科学版), 2019, 47: 103-9
- [13] 樊起傅, 付枉. 端粒与着丝粒--染色体上的高度重复序列区域. 科学通报, 2017, 62: 3245-55
- [14] Hartley G, O'Neill RJ. Centromere repeats: hidden gems of the genome. Genes (Basel), 2019, 10: 223
- [15] Talbert PB, Masuelli R, Tyagi AP, et al. Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. Plant Cell, 2002, 14: 1053-66
- [16] Alfenito MR, Birchler JA. Molecular characterization of a

- maize B chromosome centric sequence. *Genetics*, 1993, 135: 589-97
- [17] Lim KB, De Jong H, Yang TJ, et al. Characterization of rDNAs and tandem repeats in heterochromatin of *Brassica rapa*. *Mol Cells*, 2005, 19: 436-44
- [18] Gill N, Findley S, Walling JG, et al. Molecular and chromosomal evidence for allopolyploidy in soybean. *Plant Physiol*, 2009, 151: 1167-74
- [19] Gong Z, Wu Y, Koblizková A, et al. Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *Plant Cell*, 2012, 24: 3559-74
- [20] Su H, Liu Y, Liu C, et al. Centromere satellite repeats have undergone rapid changes in polyploid wheat subgenomes. *Plant Cell*, 2019, 31: 2035-51
- [21] Kamm A, Galasso I, Schmidt T, et al. Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationship between *Arabidopsis* species. *Plant Mol Biol*, 1995, 27: 853-62
- [22] Cheng ZJ, Murata M. A centromeric tandem repeat family originating from a part of *Ty3/gypsy*-retroelement in wheat and its relatives. *Genetics*, 2003, 164: 665-72
- [23] Kawabe A, Nasuda S. Structure and genomic organization of centromeric repeats in *Arabidopsis* species. *Mol Genet Genomics*, 2005, 272: 593-602
- [24] Samoluk SS, Robledo G, Bertoli D, et al. Evolutionary dynamics of an at-rich satellite DNA and its contribution to karyotype differentiation in wild diploid *Arachis* species. *Mol Genet Genomics*, 2017, 292: 283-96
- [25] Tek AL, Kashihara K, Murata M, et al. Functional centromeres in *Astragalus sinicus* include a compact centromere-specific histone H3 and a 20-bp tandem repeat. *Chromosome Res*, 2011, 19: 969-78
- [26] Gindullis F, Desel C, Galasso I, et al. The large-scale organization of the centromeric region in *Beta* species. *Genome Res*, 2001, 11: 253-65
- [27] Wen R, Moore G, Shaw PJ. Centromeres cluster *de novo* at the beginning of meiosis in *Brachypodium distachyon*. *PLoS One*, 2012, 7: e44681
- [28] Xia X, Selvaraj G, Bertrand H. Structure and evolution of a highly repetitive DNA sequence from *Brassica napus*. *Plant Mol Biol*, 1993, 21: 213-24
- [29] Harrison GE, Heslop-Harrison JS. Centromeric repetitive DNA sequences in the genus *Brassica*. *Theor Appl Genet*, 1995, 90: 157-65
- [30] Han Y, Wang G, Liu J, et al. Divergence in centromere structure distinguishes related genomes in *Coix lacryma-jobi* and its wild relative. *Chromosoma*, 2010, 119: 89-98
- [31] Kulikova O, Geurts R, Lamine M, et al. Satellite repeats in the functional centromere and pericentromeric heterochromatin of *Medicago truncatula*. *Chromosoma*, 2004, 113: 276-83
- [32] Lee HR, Neumann P, Macas J, et al. Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. *Mol Biol Evol*, 2006, 23: 2505-20
- [33] Nagaki K, Murata M. Characterization of CENH3 and centromere-associated DNA sequences in sugarcane. *Chromosome Res*, 2005, 13: 195-203
- [34] Kishii M, Nagaki K, Tsujimoto H. A tandem repetitive sequence located in the centromeric region of common wheat (*Triticum aestivum*) chromosomes. *Chromosome Res*, 2001, 9: 417-28
- [35] Malik HS, Henikoff S. Conflict begets complexity: The evolution of centromeres. *Curr Opin Genet Dev*, 2002, 12: 711-8
- [36] Presting GG. Centromeric retrotransposons and centromere function. *Curr Opin Genet Dev*, 2018, 49: 79-84
- [37] 左胜. 甘蔗着丝粒 DNA 序列组成及进化分析[D]. 福州: 福建农林大学, 2018
- [38] Feng C, Liu Y, Su H, et al. Recent advances in plant centromere biology. *Sci China Life Sci*, 2015, 58: 240-5
- [39] Li B, Choulet F, Heng Y, et al. Wheat centromeric retrotransposons: the new ones take a major role in centromeric structure. *Plant J*, 2013, 73: 952-65
- [40] Zhu Z, Gui S, Jin J, et al. The NnCENH3 protein and centromeric DNA sequence profiles of *Sacred Lotus* reveal the DNA structures and dynamics of centromeres in basal eudicots. *Plant J*, 2016, 87: 568-82
- [41] Neumann P, Navrátilová A, Koblížková A, et al. Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob DNA*, 2011, 2: 4
- [42] Pélissier T, Tutois S, Deragon JM, et al. Athila, a new retroelement from *Arabidopsis thaliana*. *Plant Mol Biol*, 1995, 29: 441-52
- [43] 余朝文, 宋运淳. 植物着丝粒结构和功能的研究进展. *遗传*, 2006, 28: 1597-606
- [44] Han J, Masonbrink RE, Shan W, et al. Rapid proliferation and nucleolar organizer targeting centromeric retrotransposons in cotton. *Plant J*, 2016, 88: 992-1005
- [45] Houben A, Schroeder-Reiter E, Nagaki K, et al. CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley. *Chromosoma*, 2007, 116: 275-83
- [46] Dong F, Miller JT, Jackson SA, et al. Rice (*Oryza sativa*) centromeric regions consist of complex DNA. *Proc Natl Acad Sci USA*, 1998, 95: 8135-40
- [47] Francki M. Identification of Bilby, a diverged centromeric *Ty1-copia* retrotransposon family from cereal rye (*Secale cereale* L.). *Genome*, 2001, 44: 266-74
- [48] Miller JT, Dong F, Jackson SA, et al. Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics*, 1998, 150: 1615-23
- [49] Saunders VA, Houben A. The pericentromeric heterochromatin of the grass *Zingiberia Biebersteiniana* (2n = 4) is composed of Zbcen1-type tandem repeats that are intermingled with accumulated dispersedly organized sequences. *Genome*, 2001, 44: 955-61
- [50] Zhong CX, Marshall JB, Topp C, et al. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell*, 2002, 14: 2825-36
- [51] Gent JI, Madzima TF, Bader R, et al. Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-directed DNA methylation. *Plant Cell*, 2014, 26: 4903-17
- [52] 欧阳凯. 芸薹属植物功能着丝粒序列组成及进化分析

- [D]. 福州: 福建农林大学, 2020
- [53] Copenhaver GP, Nickel K, Kuromori T, et al. Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science*, 1999, 286: 2468-74
- [54] Nagaki K, Cheng ZK, Ouyang S, et al. Sequencing of a rice centromere uncovers active genes. *Nat Genet*, 2004, 36: 138-45
- [55] Harkess A, Huang K, van der Hulst R, et al. Sex determination by two Y-linked genes in garden asparagus. *Plant Cell*, 2020, 32: 1790-6
- [56] You Q, Zhai K, Yang DWY, et al. An E3 ubiquitin ligase-BAG protein module controls plant innate immunity and broad-spectrum disease resistance. *Cell Host Microbe*, 2016, 20: 758-69
- [57] 李鹏. 川桑CENH3基因的鉴定及着丝粒序列初探[D]. 重庆: 西南大学, 2020
- [58] Hara M, Fukagawa T. Where is the right path heading from the centromere to spindle microtubules? *Cell Cycle*, 2019, 18: 1199-211
- [59] Wang S, Jin W, Wang K. Centromere histone H3- and phospholipase-mediated haploid induction in plants. *Plant Methods*, 2019, 15: 42
- [60] Karimi-Ashtiyani R. Centromere engineering as an emerging tool for haploid plant production: advances and challenges. *Methods Mol Biol*, 2021, 2289: 3-22
- [61] Watts A, Kumar V, Bhat SR. Centromeric histone H3 protein: from basic study to plant breeding applications. *J Plant Biochem. Biotechnol*, 2016, 25: 339-48
- [62] Kuo Y, Ishii T, Fuchs J, et al. The evolutionary dynamics of repetitive DNA and its impact on the genome diversification in the genus *Sorghum*. *Front Plant Sci*, 2021, 12: 279734
- [63] Kursel LE, Malik HS. The cellular mechanisms and consequences of centromere drive. *Curr Opin Cell Biol*, 2018, 52: 58-65
- [64] Krátká M, Smerda J, Lojdrová K, et al. Holocentric chromosomes probably do not prevent centromere drive in *Cyperaceae*. *Front Plant Sci*, 2021, 12: 642661
- [65] 丁戈, 姚南, 吴琼, 等. 着丝粒结构与功能研究的新进展. *植物学报*, 2008, 25: 149-60
- [66] Ugarković D, Plohl M. Variation in satellite DNA profiles: causes and effects. *EMBO J*, 2002, 21: 5955-9
- [67] Sharma A, Wolfgruber TK, Presting GG. Tandem repeats derived from centromeric retrotransposons. *BMC Genomics*, 2013, 14: 142
- [68] Uchida W, Matsunaga S, Sugiyama R, et al. Interstitial telomere-like repeats in the *Arabidopsis thaliana* genome. *Genes Geneti Syst*, 2002, 77: 63-7
- [69] Gong X, Hong D, Chen S, et al. Isolation and characterization of five rice telomere-associated sequences. *Sci China C Life Sci*, 1998, 41: 372-80
- [70] Tek AL, Jiang J. The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. *Chromosoma*, 2004, 113: 77-83
- [71] Villasante A, Abad JP, Mendez-Lago M. Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proc Natl Acad Sci USA*, 2007, 104: 10542-7
- [72] Ling YH, Lin Z, Yuen KW. Genetic and epigenetic effects on centromere establishment. *Chromosoma*, 2020, 129: 1-24
- [73] Koo DH, Han F, Birchler JA, et al. Distinct DNA methylation patterns associated with active and inactive centromeres of the maize B chromosome. *Genome Res*, 2011, 21: 908-14
- [74] 刘亚林, 苏汉东, 韩方普. 玉米新着丝粒形成及其表观遗传学. *遗传*, 2015, 37: 407-8
- [75] Wong CYY, Lee BCH, Yuen KKY. Epigenetic regulation of centromere function. *Cell Mol Life Sci*, 2020, 77: 2899-917
- [76] Comai L, Maheshwari S, Marimuthu MPA. Plant centromeres. *Curr Opin Plant Biol*, 2017, 36: 158-67
- [77] Carty BL, Dunleavy EM. Truly epigenetic: a centromere finds a “neo” home. *J Cell Biol*, 2021, 220: e202101027
- [78] Naughton C, Gilbert N. Centromere chromatin structure - lessons from neocentromeres. *Exp Cell Res*, 2020, 389: 111899
- [79] 王圣. 川桑着丝粒与端粒重复序列及染色体定位[D]. 重庆: 西南大学, 2016
- [80] Perea-Resca C, Blower MD. Centromere biology: transcription goes on stage. *Mol Cell Biol*, 2018, 38: e00263-18
- [81] Hill A, Bloom K. Genetic manipulation of centromere function. *Mol Cell Biol*, 1987, 7: 2397-405
- [82] Nakano M, Cardinale S, Noskov VN, et al. Inactivation of a human kinetochore by specific targeting of chromatin modifiers. *Dev Cell*, 2008, 14: 507-22
- [83] Cohen AK, Huh TY, Helleiner CW. Transcription of satellite DNA in mouse L-cells. *Can J Biochem*, 1973, 51: 529-32
- [84] Liu Q, Liu Y, Shi Q, et al. Emerging roles of centromeric RNAs in centromere formation and function. *Genes Genomics*, 2021, 43: 217-26
- [85] Corless S, Höcker S, Erhardt S. Centromeric RNA and its function at and beyond centromeric chromatin. *J Mol Biol*, 2020, 432: 4257-69
- [86] Talbert PB, Henikoff S. Transcribing centromeres: noncoding RNAs and kinetochore assembly. *Trends Genet*, 2018, 34: 587-99
- [87] Topp CN, Zhong CX, Dawe RK. Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc Natl Acad Sci USA*, 2004, 101: 45: 15986-91
- [88] Chan FL, Marshall OJ, Saffery R, et al. Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proc Natl Acad Sci USA*, 2012, 109: 1979-84
- [89] 刘青. 植物着丝粒结构及进化的研究进展. *热带亚热带植物学报*, 2015, 23: 576-86
- [90] Foltz DR, Jansen LE, Black BE, et al. The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol*, 2006, 8: 458-69
- [91] Ishii T, Juranić M, Maheshwari S, et al. Unequal contribution of two paralogous CENH3 variants in cowpea centromere function. *Commun Biol*, 2020, 3: 775
- [92] Feng C, Yuan J, Bai H, et al. The deposition of CENH3 in

- maize is stringently regulated. *Plant J*, 2020, 102: 6-17
- [93] Maheshwari S, Tan EH, West A, et al. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet*, 2015, 11: e1004970
- [94] Kelliher T, Starr D, Wang W, et al. Maternal haploids are preferentially induced by CENH3-tailswap transgenic complementation in maize. *Front Plant Sci*, 2016, 7: 414
- [95] Ravi M, Chan SW. Haploid plants produced by centromere-mediated genome elimination. *Nature*, 2010, 464: 615-8
- [96] 吴琼, 程志号, 孙长君, 等. 着丝粒DNA序列的分离和筛选方法研究进展. *分子植物育种*, 2021: 1-18
- [97] Han M, Yang Y, Zhang M, et al. Considerations regarding centromere assembly in plant whole-genome sequencing. *Methods*, 2021, 187: 54-6
- [98] 薛超, 张融, 郭瑞, 等. 水稻着丝粒特异组蛋白CENH3抗体的制备与应用. *中国水稻科学*, 2017, 31: 475-82
- [99] Novák P, Neumann P, Pech J, et al. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, 2013, 29: 792-3
- [100] Neumann P, Navrátilová A, Schroeder-Reiter E, et al. Stretching the rules: monocentric chromosomes with multiple centromere domains. *PLoS Genet*, 2012, 8: e1002777
- [101] Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 2009, 10: 669-80
- [102] Xia QM, Miao LK, Xie KD, et al. Localization and characterization of *Citrus* centromeres by combining half-tetrad analysis and CENH3-associated sequence profiling. *Plant Cell Rep*, 2020, 39: 1609-22
- [103] Song JM, Xie WZ, Wang S, et al. Two gap-free reference genomes and a global view of the centromere architecture in rice. *Mol Plant*, 2021, 14: 1757-67
- [104] Naish M, Alonge M, Wlodzimierz P, et al. The genetic and epigenetic landscape of the *Arabidopsis* centromeres. *Science*, 2021, 374: eabi7489
- [105] Vondrak T, Robledillo LA, Novák P, et al. Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. *Plant J*, 2020, 101: 484-500
- [106] Yang Z, Ge X, Li W, et al. Cotton D genome assemblies built with long-read data unveil mechanisms of centromere evolution and stress tolerance divergence. *BMC Biol*, 2021, 19: 115
- [107] Kirov I, Odincov S, Omarov M, et al. Functional *Allium fistulosum* centromeres comprise arrays of a long satellite repeat, insertions of retrotransposons and chloroplast DNA. *Front Plant Sci*, 2020, 11: 562001
- [108] Balzano E, Giunta S. Centromeres under pressure: evolutionary innovation in conflict with conserved function. *Genes (Basel)*, 2020, 11: 912
- [109] Oliveira LC, Torres GA. Plant centromeres: genetics, epigenetics and evolution. *Mol Biol Rep*, 2018, 45: 1491-7
- [110] Ohzeki J, Larionov V, Earnshaw WC, et al. *De novo* formation and epigenetic maintenance of centromere chromatin. *Curr Opin Cell Biol*, 2019, 58: 15-25
- [111] Zhao J, Hao W, Tang C, et al. Plasticity in *Triticeae* centromere DNA sequences: a wheat × tall wheatgrass (decaploid) model. *Plant J*, 2019, 100: 314-27
- [112] Liu Y, Su H, Zhang J, et al. Rapid birth or death of centromeres on fragmented chromosomes in maize. *Plant Cell*, 2020, 32: 3113-23
- [113] Huang Y, Ding W, Zhang M, et al. The formation and evolution of centromeric satellite repeats in *Saccharum* species. *Plant J*, 2021, 106: 616-29