生物技术进展 2022 年 第 12 卷 第 1 期 1~9
Current Biotechnology ISSN 2095-2341

进展评述
Reviews

植物功能着丝粒DNA研究进展

张会[1,2]， 王凯[1*]

1.南通大学生命科学学院，江苏 南通 226019；
2.南通大学教务处，江苏 南通 226019

摘 要： 着丝粒(centromere)是真核生物染色体的重要功能结构。在细胞有丝分裂和减数分裂过程中，着丝粒通过招募动粒蛋白行使功能，保障染色体正确分离和传递。真核生物中，含有着丝粒特异组蛋白的CenH3区域被定义为功能着丝粒区，即真正意义上的着丝粒。近年来，借助染色质免疫沉淀技术，人们对功能着丝粒DNA开展了深入研究，揭示其组成、结构及演化特征，并发现功能着丝粒区存在具有转录活性的基因，且部分基因具有重要生物学功能。由于存在大量重复DNA，着丝粒演化之谜一直未能完全揭示。对植物功能着丝粒DNA序列研究进展进行了概述，并重点阐述了着丝粒重复DNA研究的新方法和新进展，以期为深入开展相关研究提供借鉴。

关键词：功能着丝粒；CenH3；重复DNA；演化

DOI：10.19586/j.2095-2341.2021.0109

中图分类号：Q343.2 **文献标志码：**A

The Research Progress on Plant Functional Centromere DNAs

ZHANG Hui[1,2] , WANG Kai[1*]

1.*School of Life Science*, *Nantong University*, *Jiangsu Nantong* 226019, *China*;
2.*Academic Affairs Office*, *Nantong University*, *Jiangsu Nantong* 226019, *China*

Abstract: The centromere is an important functional region in eukaryotic chromosomes. During cell mitosis and meiosis in eukaryotes, centromere plays a key role in faithful segregation of chromosomes by recruiting the kinetochore. In eukaryotes, the CenH3 region containing centromere specific histones is defined as the functional centromere region, which is the true centromere. In recent years, based on the application of chromatin immunoprecipitation, centromeric DNAs can be readily isolated and characterized. The composition, construction and evolution of centromeres were revealed, and some genes were found with important biological functions. However, because of the high level of diversity of centromeric repetitive DNAs, the evolution of centromeric repetitive DNAs is still elusive. In this paper, the research progress of functional centromere DNA sequences in plants was summarized, and the new methods and progress of centromere repetitive DNAs research were emphasized, which aimed to provide reference for further research.

Key words: functional centromere; CenH3; repetitive DNAs; evolution

着丝粒是染色体三大功能元件之一，在真核生物细胞分裂过程中发挥着保障染色体正确分离的重要作用。着丝粒在染色体形态上表现为中期的主缢痕，分子结构上则是一个DNA-蛋白的复合体。具体而言，着丝粒主要由内部的着丝粒染色质(DNA及组蛋白)以及外部的蛋白复合体组成。这一外部复合体即动粒(kinetochore)结构，在细胞分裂过程中，纺锤丝微管通过与动粒相连从而实现牵引染色体的运动与分离(图1)[1-2]。目前，已发现的动粒蛋白超过100种，且在不同物种间都显示出较高的保守性[3-6]。

着丝粒的显著特征之一是其核小体含有H3

收稿日期：2021-06-15；**接受日期：**2021-08-05
基金项目：国家自然科学基金项目(32070544)。
联系方式：张会 E-mail：hzhangfly107@126.com； *通信作者 王凯 E-mail：kwang5@126.com

组蛋白变体CenH3(果蝇中称作CID,动物中称作CENP-A)[7]。CenH3与H3组蛋白虽然具有高度相似性,但不同物种在其N端存在高度变异,即使在近缘物种间仍可能存在一定差异。因此,CenH3的N端序列常用于开发物种特异引物和制备CenH3抗体。CenH3的组蛋白折叠区(histone fold domain,HFD)是另一个重要区域,其内部的第1个折叠环(loop1)和第2个α折叠螺旋(α2-helix)在CenH3靶向结合到着丝粒核小体过程中发挥决定性作用。因此,这一区域被称为CenH3的靶向结构域(CenH3-targeting domain,CATD)[3,5]。另外,基于CenH3在真核生物着丝粒中的普遍存在特征以及在动粒组装过程的重要作用[8],CenH3也被作为功能着丝粒的标志[9-12],用于界定真正意义上的着丝粒区。

着丝粒区域往往由大量重复序列DNA组成,主要包含串联重复序列(亦称卫星重复序列,tandem repeat or satellite repeat)与转座子重复序列。虽然着丝粒功能及其组成蛋白在物种间高度保守,但是重复DNA序列则呈现快速演化特征,具有高度物种特异性。近年来,染色质免疫共沉淀结合高通量测序技术(chromatin immunocoprecipitation combined with high-throughput sequencing,ChIP-seq)以及重复序列分析等新技术解决了着丝粒DNA分离的难题,从而推动了着丝粒DNA结构及演化研究的深入开展。另外,作为染色体的基本功能元件之一,着丝粒也是合成生物学研究的重要内容。本文针对植物着丝粒DNA的相关研究进行了综述,以期为后继研究提供借鉴。

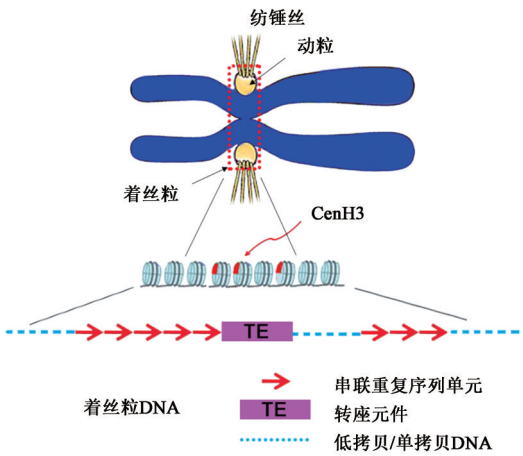


图1 着丝粒结构

Fig. 1 The structure of centromere

1 着丝粒DNA

植物着丝粒DNA主要由串联重复亦称卫星重复序列、反转录转座子(centromeric retrotransposon)以及低拷贝序列组成[13-15]。串联重复序列由拷贝数不等的重复单元组成[16-17],往往与着丝粒特异的反转录转座子交替分布于着丝粒区[18-19]。着丝粒区同时存在低拷贝序列,并包含具有转录活性的基因[20-22]。

1.1 着丝粒串联重复序列

真核生物中,串联重复是着丝粒区域分布最为丰富的一类重复序列。研究表明,着丝粒串联重复在某些物种中可以占据几乎全部着丝粒区域,如人类中,由171 bp重复单元组成的卫星重复(亦称α卫星DNA)可组成长度250 kb~5 Mb的串联序列,占据着丝粒的绝大部分区域,并可延伸至功能着丝粒区外部。串联重复也是多数植物着丝粒的主要成分(表1),如拟南芥(*Arabidopsis thaliana*)着丝粒长度范围在2.8~4.0 Mb之间,其主要由重复单元为180 bp的串联重复组成;二穗短柄草(*Brachypodium distachyon*)、水稻(*Oryza sativa*)、玉米(*Zea mays*)以及甘蔗(*Saccharum officinarum*)等禾本科植物的着丝粒同样含有大量串联重复序列,这些串联重复的重复单元大小为140~156 bp,可组成长度达2 Mb的序列,占据着丝粒的大部分区域[22, 24-26]。

研究发现,不同物种着丝粒串联重复的显著特征是重复单元大小近似于缠绕1个核小体的DNA长度,即约150 bp。分析认为,这一特征可能与着丝粒核小体特殊的结构及DNA缠绕方式有关。然而,在马铃薯与甘蔗的研究中发现,其着丝粒串联重复的重复单元可长达数kb[19,22]。分析证实,这些重复序列与反转录转座子具有高度相似性,说明其很可能来源于反转录转座子。推测反转录转座子的频繁转座事件推动了具有长重复单元的串联重复序列的形成,而这些长重复单元的串联重复经适应性选择后演变成具有约150 bp的着丝粒串联重复单元。由此认为,那些具有重复单元约为150 bp串联重复序列的着丝粒可能处于进化上的中后期,即处于成熟着丝粒阶段;反之,长重复单元的着丝粒则可能处于进化上相对较早的阶段。

表1 植物中已知的着丝粒串联重复序列[23]

Table 1 Known centromeric tandem repeats in plants[23]

属名	种名	重复序列	序列大小/bp
拟南芥属（*Arabidopsis*）	大沙叶芥（*A. arenosa*）	pAa214	166~179
	叶芽鼠耳芥（*A. gemmifera*）	pAge1，pAge2	180
	无苞芥（*A. pumila*）	pAgKB2，pAgKB1	180
	鼠耳芥（*A. thaliana*）	pAL1	180
落花生属（*Arachis*）	*A. batizocoi*	AbaTR-2	297
	落花生（*A. hypogaea*）	Cot-1-5，Cot-1-59，Cot-1-119	115
	A. helodes	AheTR-2	295
	A. trinitensis	AtrTR-2	300
甜菜属（*Beta*）	白花甜菜（*B. corolliflora*）	pHC8	162
	平匍甜菜（*B. procumbens*）	pTS5，pTS4.1	158~312
	栽培甜菜（*B. vulgaris*）	pBv1	326~327
芸薹属（*Brassica*）	紫菜薹（*B. campestris*）	pBcKB4，pBT11	175
	结球甘蓝（*B. oleracea*）	pBoKB1	171
稻属（*Oryza*）	短花药野生稻（*O. brachyantha*）	CentO-F	154
	根茎野生稻（*O. rhizomatis*）	CentO-C1，CentO-C2	126，366
	栽培稻（*O. sativa*）	CentO	155
茄属（*Solanum*）	*S. bulbocastanum*	pSbTC1	7
	马铃薯（*S. tuberosum*）	St18，St24，St49，St57，St3-58，St3-238，St3-294	979~5 390
	紫云英（*Astragalus sinicus*）	CentAs	20
	二穗短柄草（*Brachypodium distachyon*）	CentBd	156
	大豆[*Glycine max* (soybean)]	CentGm-1，CentGm-2，CentGm-4	91~411
	光叶百脉根（*Lotus japonicus*）	Ljcen1	440
	紫苜蓿（*Medicago sativa*）	MsCR	15~168
	御谷（*Pennisetum glaucum*）	pPgKB19	137
	甘蔗热带种（*Saccharum officinarum*）	SCEN	140
	高粱[*Sorghum bicolor* (sorghum)]	pSau3A10	137
	小麦[*Triticum aestivum* (wheat)]	Tail family	570
	三叶草（*Trifolium repens*）	TrR350	350
	豇豆（*Vigna unguiculata*）	pVuKB1	488
	玉米（*Zea mays*）	CentC,Cent4	156，740

着丝粒串联重复序列通常具有种属高度特异性，不同物种间存在极低相似性，这说明着丝粒串联重复序列具有快速演化特性[6]。不同物种比较发现，着丝粒串联重复序列具有极低的同源性[2]，即使是属内亲缘相近的种间材料，其着丝粒串联重复序列也可能不尽相同，如栽培稻的着丝粒串联重复序列CentO在稻属（*Oryza*）中广泛存在，但在近缘的FF组野生稻（*Oryza brachyantha*）和CC组野生稻（*Oryza rhizomatis*）的着丝粒中却并未发现[19, 21]。目前，仅发现玉米、水稻等禾本科物种的着丝粒串联重复单元间具有约80 bp的同源区段[27-29]，这一保守的序列被认为包含了着丝粒核心功能信息。深入分析发现，这一保守序列可形成与着丝粒核小体间更佳的缠绕结构，这一点也说明了着丝粒DNA序列的结构在其功能行使中发挥重要作用[22]。

1.2 着丝粒反转录转座子序列

着丝粒反转录转座子（centromeric retrotransposons）是分布于着丝粒区域的一类转座元件，往往为LTR类型反转录转座子[30]。研究发现，植物着丝粒的转座元件大部分是由LTR型的Ty3/gypsy类反转录转座子组成[31-32]（表2）。例如，拟南

芥着丝粒特异的反转录转座子Athila(Ty3/gypsy)与串联重复穿插分布于着丝粒,两者共同形成着丝粒核心区域3 Mb的着丝粒中央结构域;与之类似,水稻着丝粒也由大量Ty3/gypsy反转录转座子CRR(centromere-specific retrotransposon of rice)与着丝粒串联重复CentO组成[27, 33]。同样,着丝粒特异的反转录转座子CRM(CR of maize)也大量分布在玉米着丝粒中[34]。已知最古老的被子植物无油樟的研究表明,转座元件在其进化早期就已经插入到着丝粒中[35],说明着丝粒反转录转座子在早期着丝粒形成中发挥重要作用。

表2 植物着丝粒特异反转录转座子序列[23]

Table **2** The centromeric retrotransposons in plants[23]

种名	重复序列	序列大小/bp
拟南芥(*Arabidopsis thaliana*)	Athila	10 500
落花生(*Arachis hypogaea*)	Cot-1-Clone117(partial)	117
平匍甜菜(*Beta procumbens*)	pBp10	417
甜菜(*Beta vulgaris*)	pBv26	417
二穗短柄草(*Brachypodium distachyon*)	CRBd1	—
大豆[*Glycine max* (soybean)]	GmCR	421~852
雷蒙德棉(*Gossypium raimondii*)	Gr087,Gr201,Gr285,Gr291,Gr298,Gr304,Gr327,Gr334,Gr344,Gr359,Gr369,Gr372,Gr386,Gr407,Gr442,Gr452,Gr471,Gr552,Gr616	309~5 909
大麦(*Hordeum vulgare*)	cereba	7 176
番茄(*Lycopersicum esculentum*)	TGRIV	7 000
水稻(*Oryza sativa*)	CRR	7 400~7 800
黑麦(*Secale cereale*)	Bilby	3 400
甘蔗热带种(*Saccharum officinarum*)	CRS	3 600
高粱(*Sorghum bicolor*)	pSau3A9	745
野生一粒小麦(*Triticum boeoticum*)	CRWs	7 762~7 865
玉米(*Zea mays*)	CRM	7 572
Zingeria biebersteiniana	Zb47A	263

着丝粒反转录转座子序列的分布虽然具有一定的着丝粒特异性,但往往也会延伸分布于着丝粒旁侧区域。例如,对玉米、水稻、棉花等研究发现,着丝粒反转录转座子序列主要集中在具有CenH3的功能着丝粒区域,但是,也有大量拷贝延伸分布到非CenH3结合的着丝粒附近区域,甚至在远离着丝粒的染色体臂上亦有发现[36];同时,棉花着丝粒反转录转座子也被发现与rDNA分布重叠,暗示着丝粒与rDNA重复序列间可能存在某种共有特性,适于着丝粒反转录转座子的插入[37]。这一发现为探究着丝粒反转录转座子的形成与着丝粒区特异富集机制提供了新的线索。

然而,并非所有植物着丝粒LTR类反转录转座子都是Ty3/gypsy类型。在中国古代莲(*Nelumbo nucifera* Gaertn.)的研究中发现,其着丝粒主要分布于Ty1/copia类反转录转座子上[38];另外,Ty1/copia类型反转录转座子也是小麦着丝粒的主要转座元件[39]。值得注意的是,古代莲和小麦着丝粒中仍然存在部分Ty3/gypsy类型的反转录转座子序列[38],表明这两类反转录转座子均可形成着丝粒特异序列。但是,Ty3/gypsy类型的反转录转座子序列如何在多数物种着丝粒中获得进化优势仍然是未解之谜。

1.3 功能基因序列

由于存在大量重复序列以及异染色质化,着丝粒区遗传重组及转录活性受到显著抑制,早期

认为不存在功能基因[40-41]。水稻8号染色体着丝粒中具有转录活性基因的发现突破了这一认知[10,42]。目前，在模式材料拟南芥、二穗短柄草以及马铃薯、棉花等作物着丝粒区均发现存在转录活性基因[17,19,21,24,37]。但与常染色质区不同，着丝粒区的基因密度显著低于常染色质区，例如，棉花着丝粒的基因密度为510 kb分布1个基因，而基因组其他区域的密度则达到20 kb分布1个基因[37]；同样，在禾本科模式物种二穗短柄草的着丝粒区平均每60 kb分布1个基因，而在非着丝粒区基因的密度为8 kb分布1个基因[24]；并且，这些基因往往分布在H3组蛋白存在的核小体区域，CenH3核小体区域则极少含有转录活性基因，这一点也说明CenH3的存在与基因转录具有负相关的特征。目前，着丝粒区的基因表达还未发现具有普遍的时空特异性，但其表达量通常较低[24,37,43]，这也印证了着丝粒抑制基因表达的特性。

分析发现，着丝粒基因功能多数是未知的。但在水稻、木瓜中的研究揭示着丝粒中的基因可能具有重要的生物学意义。抑制重组是性染色体形成的主要因素，交换抑制导致性别决定基因不能被交换到同源染色体上，最终将性别决定基因固定到特定染色体，从而形成了带有固定性别决定基因的性染色体。在木瓜的研究中发现，其性别决定区位于着丝粒内[44]，这与上述着丝粒区抑制交换从而形成性染色体的猜想是吻合的。水稻中，*EBR1*是一个介导广谱细菌和真菌抗性的基因，而该基因也被定位于着丝粒区[45]。印证着丝粒可能存在具有重要生物学功能的基因。这一基因的定位和克隆经过了长期的研究，这主要是由于着丝粒区交换频率低、难以获得重组个体导致，这一点值得研究者在开展着丝粒区基因的研究中加以关注。总之，木瓜性别决定基因和水稻*EBR1*基因的研究证实了着丝粒区可能存在具有特殊生物学功能的基因，但其定位和克隆将是一个挑战。

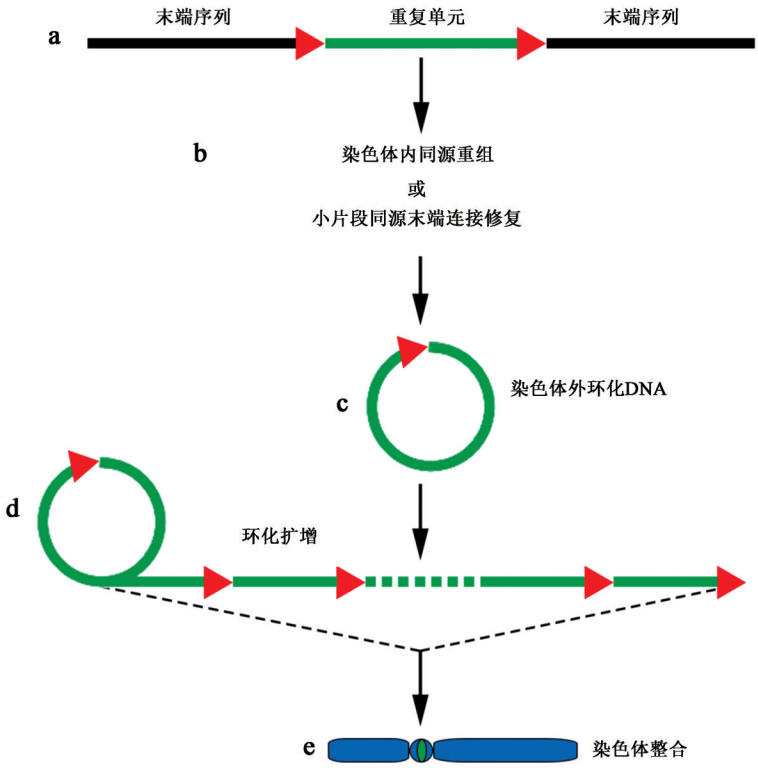
2 着丝粒DNA进化机制

研究认为，着丝粒起源于新着丝粒区，与成熟着丝粒不同，新着丝粒位于缺乏重复序列的常染色质区，但随着重复序列在新着丝粒的产生和扩增，最后占据功能着丝粒区域，形成富含串联重复的成熟着丝粒[21]。然而，着丝粒重复序列如何形成并特异富集在着丝粒区仍然未有定论。目前认为，着丝粒形成后，可能存在多种不同的串联重复序列，并在未知的选择压力下某一类型形成适应着丝粒结构后迅速占据主导地位，从而成为着丝粒区的主要成分[19,21]，这一观点在马铃薯着丝粒的研究中得以验证。马铃薯12个着丝粒分布着不同类型的串联序列，并且有5条染色体的着丝粒区并未发现串联重复序列存在；同时，比较分析发现，多个串联重复序列在马铃薯近缘物种的着丝粒中并不存在，进而说明这些序列处于快速演化的阶段。但这些结果仍然无法解答着丝粒串联序列如何形成、如何在着丝粒中特异富集等问题。

有研究认为，着丝粒反转录转座子通过其高频转座形成串联重复序列，但这一假设一直未能得到证实。在对陆地棉祖先种雷蒙德棉的研究中发现，其着丝粒中仅含有反转录转座子序列，并未发现卫星序列存在；同时，反转录转座子序列也呈现多种类型[37]。这一发现暗示，着丝粒重复序列可能起源于反转录转座子。而在马铃薯、甘蔗等研究中的发现为这一观点提供了新的证据。马铃薯与甘蔗着丝粒均存在长重复单元的串联重复序列，而其序列与反转录转座子呈现高度相似性，由此可以推断，这些串联重复序列很可能来自反转录转座子[19,21-22]。最近，在甘蔗中发现了多个着丝粒串联重复序列具有末端重复序列特征；分析显示，这一结构可导致DNA发生环化[22]，而通过DNA的环化扩增，可以快速形成不同长度的串联重复序列[46]；这些串联重复序列可以借助末端同源序列重新整合到染色体上（图2），从而为着丝粒反转录转座子形成串联重复序列提供了新的线索，也为着丝粒串联序列演化提出了一个新的完整的路径。

着丝粒重复序列如何特异地在着丝粒富集一直是一个未解之谜。通过对着丝粒重复序列与端粒序列比较分析发现，在少数物种中着丝粒序列与端粒序列具有较高的相似性[47-49]。同样，在棉花中的研究发现，着丝粒反转录转座子序列与rDNA在染色体上呈现重叠分布，比较分析并未发现两者具有任何序列同源性。但这一发现可能暗示丝粒区与rDNA及端粒区域存在某种类似结构特征，导致了着丝粒反转录转座子的特异插入[37]。

这一结构特征可能是着丝粒核小体的特殊结构所决定的。例如，在水稻、柳枝稷及甘蔗中均发现其占主要成分的重复序列与CenH3核小体具有一定缠绕吻合特征，这一特征使得着丝粒重复序列与核小体形成更加牢固的结构，利于着丝粒稳定结构和行使功能[22,29,50]。这些发现进一步说明，在着丝粒重复序列的演化过程中，具有适宜着丝粒核小体结构的重复序列可能发挥了决定性的选择作用，推动了着丝粒重复序列的演化以及着丝粒特异富集。



注：含有末端重复序列（红色三角形）的着丝粒反转录转座子（a）通过染色体内同源重组或者小片段同源末端连接修复等机制（b）形成游离于染色体外的环化DNA（c）；环化DNA可经过环化扩增，形成不同长度串联重复序列（d）；并通过同源重组等机制再整合到染色体上的着丝粒区（e）。

图2 基于体外环化机制的着丝粒串联重复DNA演化模式

Fig 2 The proposed eccDNA-based on rolling circle amplification mechanism.

3 着丝粒DNA序列的分离与分析方法

由于着丝粒重复序列的物种特异性，常规比较基因组学方法往往难以对其进行有效的分离和分析。近年来，根据着丝粒存在特异组蛋白CenH3的特点，研究者建立了基于CenH3抗体的ChIP-seq的着丝粒分离分析方法[51]，使得分离不同物种着丝粒DNA得以实现，并极大推动了着丝粒研究的深入开展。

在获得着丝粒ChIP-seq数据后，通过与基因组序列比对筛选，获得在基因组上具有单一拷贝的数据，并将其锚定到对应物种的基因组上，即可获得ChIP-seq富集的峰图，该峰图则指示着丝粒区。这一方法目前已被用于着丝粒的基因组精确定位[21,24,37,52]；同时，由于大量高度重复序列的存在，着丝粒往往是基因组拼装工作中最为困难的区域，因此，着丝粒区的组装质量往往决定和反应了基因组的整体组装质量。因此，可以根据每条染色体是否获得单一、完整的着丝粒区间来判断基因组组装质量[51]。这一点对基因组组装质量评估具有重要的参考价值。

由于着丝粒DNA的复杂性，目前已绘制基因组草图的物种极少有完整组装的着丝粒。因此，

难以将ChIP-seq数据通过与基因组比对获得着丝粒特异重复序列。近年来，由Nowak等[53]提出的1种重复序列分析方法在着丝粒的重复序列研究中广泛应用，其原理是采用较低深度的基因组数据构建重复序列家族，并根据序列相似性组装并聚类成不同的重复簇；然后将均一化后的ChIP-seq与基因组对照（input）数据锚定到不同重复簇上，从而获得两者在各个重复簇上的富集度差，具有高的ChIP-seq/input富集度差的重复序列则被认为是在着丝粒中富集的，即候选着丝粒重复序列[19,21]。结合细胞学的荧光原位杂交技术验证，最终可鉴定出着丝粒区域特异的重复序列。这一策略已在包括马铃薯、玉米、棉花、二穗短柄草、甘蔗、古代莲等多个物种的着丝粒重复序列分析中得以成功应用[21-22,24,26,37-38]，充分显示了该方法在着丝粒重复序列鉴定与分析中的有效性。随着三代单分子测序技术的发展，着丝粒序列的研究也将更加深入。结合单分子测序数据以及原有的ChIP-seq测序数据，可以更加确切地获得重复序列在着丝粒区的分布方式，进一步揭开着丝粒序列的组成与演化之谜。

4 展望

着丝粒是真核生物染色体不可或缺的功能元件，继续开展着丝粒组成、结构及演化研究是揭示着丝粒功能之谜的关键；同时，着丝粒、端粒及复制原点是合成有功能染色体的3个基本元件，因此，着丝粒形成与演化之谜的揭示也是未来合成生物学的必经之路。

就现阶段而言，尽管测序技术已有极大提升，但对于长度较长且富含重复序列的着丝粒而言，其正确拼装仍然是基因组绘制的一个巨大挑战，着丝粒区也因此被称为是基因组研究的最后阵地。着丝粒序列的鉴定和分析对于基因组组装工作而言具有重要的推动作用和借鉴意义。借助着丝粒组蛋白ChIP-seq的分析策略，实现了不同物种着丝粒序列的正确分离和鉴定，借助ChIP-seq的单拷贝序列回帖基因组实现了着丝粒的精确定位，这些极大推动了着丝粒区的组装工作。随着更长读长和更高准确率的测序技术的应用，着丝粒重复序列将被更加准确的拼接组装，着丝粒序列组成之谜也将进一步被破解，这也将推动着丝粒起源、演化乃至基因组工作的深入开展。

参 考 文 献

[1] BLOOM K S. Centromeric heterochromatin: the primordial segregation machine [J]. Annu. Rev. Genet., 2014, 48(1): 457-484.

[2] AHMAD S F, SINGCHAT W, JEHANGIR M, *et al*.. Dark matter of primate genomes: satellite DNA repeats and their evolutionary dynamics [J/OL]. Cells, 2020, 9: 2714[2020-12-24]. https://doi.org/10.3390/cells9122714.

[3] BIGGINS S. The composition, functions, and regulation of the budding yeast kinetochore [J]. Genetics, 2013, 194(4): 817-846.

[4] VARMA D, WAN X, CHEERAMBATHUR D, *et al*.. Spindle assembly checkpoint proteins are positioned close to core microtubule attachment sites at kinetochores [J]. J. Cell Biol., 2013, 202(5): 735-746.

[5] WESTHORPE F G, STRAIGHT A F. Functions of the centromere and kinetochore in chromosome segregation [J]. Curr. Opin. Cell Biol., 2013, 25(3): 334-340.

[6] PRESTING G. Centromeric retrotransposons and centromere function [J]. Curr. Opin. Genet. Dev., 2018, 49: 79-84.

[7] MALIK H S, HENIKOFF S. Major evolutionary transitions in centromere complexity [J]. Cell, 2009, 138(6): 1067-1082.

[8] LERMONTOVA I, KOROLEVA O, RUTTEN T, *et al*.. Knockdown of CENH3 in *Arabidopsis* reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation [J]. Plant J., 2011, 68(1): 40-50.

[9] TALBERT P B, MASUELLI R, TYAGI A P, *et al*.. Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant [J]. Plant Cell, 2002, 14(5): 1053-1066.

[10] NAGAKI K, CHENG Z, OUYANG S, *et al*.. Sequencing of a rice centromere uncovers active genes [J]. Nat. Genet., 2004, 36(2): 138-145.

[11] TEK A L, KASHIHARA K, MURATA M, *et al*.. Identification of the centromere-specific histone H3 variant in *Lotus japonicus* [J]. Gene, 2014, 538(1): 8-11.

[12] KIXMOELLER K, ALLU P K, BLACK B E. The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle [J/OL] . Open Biol., 2020, 10(6): 200051[2020-06-10]. http://doi.org/10.1098/rsob.200051.

[13] MADDOX P S, PORTIER N, DESAI A, *et al*.. Molecular analysis of mitotic chromosome condensation using a quantitative time-resolved fluorescence microscopy assay [J]. Proc. Natl. Acad. Sci. USA, 2006, 103(41): 15097-15102.

[14] VENKEI Z, PRZEWLOKA M R, LADAK Y, *et al*.. Spatiotemporal dynamics of Spc105 regulates the assembly of the *Drosophila* kinetochore [J/OL] . Open Biol., 2012, 2(2): 110032 [2012-02-01]. https://doi.org/10.1098/rsob.110032.

[15] OLIVEIRA L C, TORRES G A. Plant centromeres: genetics, epigenetics and evolution [J]. Mol. Biol. Rep., 2018, 45:1491-1497.

[16] MELTERS D P, BRADNAM K R, YOUNG H A, *et al*.. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution [J/OL]. Genome Biol., 2013, 14(1): R10[2013-01-30]. http://doi. org/10.1186/gb-2013-14-1-r10.

[17] COPENHAVER G P, NICKEL K, KUROMORI T, *et al*.. Genetic definition and sequence analysis of *Arabidopsis* centromeres [J]. Science, 1999, 286(5449): 2468-2474.

[18] GAO D, GILL N, KIM H R, *et al*.. A lineage-specific centromere retrotransposon in *Oryza brachyantha* [J]. Plant J., 2009, 60(5): 820-831.

[19] ZHANG H, KOBLIZKOVA A, WANG K, *et al*.. Boom-bust turnovers of megabase-sized centromeric dna in *Solanum* species: rapid evolution of dna sequences associated with centromeres [J]. Plant Cell, 2014, 26(4): 1436-1447.

[20] WU Y, KIKUCHI S, YAN H, *et al*.. Euchromatic subdomains in rice centromeres are associated with genes and transcription [J]. Plant Cell, 2011, 23(11): 4054-4064.

[21] GONG Z, WU Y, KOBLIZKOVA A, *et al*.. Repeatless and repeat-based centromeres in potato: implications for centromere evolution [J]. Plant Cell, 2012, 24(9): 3559-3574.

[22] HUANG Y, DING W, ZHANG M, *et al*.. The formation and evolution of centromeric satellite repeats in *Saccharum* species [J]. Plant J., 2021, 106(3): 616-629.

[23] 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.

[24] LI Y, ZUO S, ZHANG Z, *et al*.. Centromeric DNA characterization in the model grass *Brachypodium distachyon* provides insights on the evolution of the genus [J]. Plant J., 2018, 93(6): 1088-1101.

[25] HIATT E N, KENTNER E K, Dawe R K. Independently regulated neocentromere activity of two classes of tandem repeat arrays [J]. Plant Cell, 2002, 14(2): 407-420.

[26] ZHANG W, ZUO S, LI Z, *et al*.. Isolation and characterization of centromeric repetitive DNA sequences in *Saccharum spontaneum* [J/OL]. Sci. Rep., 2017, 7: 41659[2017-01-31]. https://doi.org/10.1038/srep41659.

[27] CHENG Z, DONG F, LANGDON T, *et al*.. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon [J]. Plant Cell, 2002, 14(8): 1691-1704.

[28] LEE H R, ZHANG W, LANGDON T, *et al*.. Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species [J]. Proc. Natl. Acad. Sci. USA, 2005, 102(33): 11793-11798.

[29] YANG X, ZHAO H, ZHANG T, *et al*.. Amplification and adaptation of centromeric repeats in polyploid switchgrass species [J]. New Phytol., 2018, 218(4): 1645-1657.

[30] LAMB J C, BIRCHLER J A. The role of DNA sequence in centromere formation [J/OL]. Genome Biol., 2003, 4(5): 214[2003-04-29]. https://doi.org/10.1186/gb-2003-4-5-214.

[31] ZHONG C X, MARSHALL J B, TOPP C, *et al*.. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3 [J]. Plant Cell, 2002, 14(11): 2825-2836.

[32] JIANG J, BIRCHLER J A, PARROTT W A, *et al*.. A molecular view of plant centromeres[J]. Trends Plant Sci., 2003, 8 (12): 570-575.

[33] FENG Q, ZHANG Y J, HAO P, *et al*.. Sequence and analysis of rice chromosome 4 [J]. Nature, 2002, 420(6913): 316-320.

[34] ANANIEV E V, PHILLIPS R L, RINES H W. Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions [J]. Proc. Natl. Acad. Sci. USA, 1998, 95(22): 13073-13078.

[35] AMBORELLA G P. The *Amborella* genome and the evolution of flowering plants [J]. Science, 2013, 342(6165): 1241089[2013-12-20]. https://www.science.org/doi/10.1126/science.

[36] JIANG J, BIRCHLER J A. Plant Centromere Biology [M]. USA: Wiley-Blackwell, 2013.

[37] HAN J, MASONBRINK R E, SHAN W, *et al*.. Rapid proliferation and nucleolar organizer targeting centromeric retrotransposons in cotton [J]. Plant J., 2016, 88(6): 992-1005.

[38] ZHU Z, GUI S, JIN J, *et al*.. The NnCenH3 protein and centromeric DNA sequence profiles of *Nelumbo nucifera* Gaertn. (*Sacred lotus*) reveal the DNA structures and dynamics of centromeres in basal eudicots[J]. Plant J., 2016, 87(6): 568-582.

[39] LI B, CHOULET F, HENG Y, *et al*.. Wheat centromeric retrotransposons: the new ones take a major role in centromeric structure [J]. Plant J., 2013, 73(6): 952-965.

[40] YAN H, JIN W, NAGAKI K, *et al*.. Transcription and histone modifications in the recombination-free region spanning a rice centromere [J]. Plant Cell, 2005, 17(12): 3227-3238.

[41] SUN X, WAHLSTROM J, KARPEN G. Molecular structure of a functional Drosophila centromere[J]. Cell, 1997, 91(7): 1007-1019.

[42] WU J, YAMAGATA H, HAYASHI-TSUGANE M, *et al*.. Composition and structure of the centromeric region of rice chromosome 8 [J]. Plant Cell, 2004, 16(4): 967-976.

[43] WANG K, WU Y, ZHANG W, *et al*.. Maize centromeres expand and adopt a uniform size in the genetic background of oat [J]. Genome Res., 2014, 24(1): 107-116.

[44] IOVENE M, YU Q, MING R, *et al*.. Evidence for emergence of sex-determining gene(s) in a centromeric region in *Vasconcellea parviflora* [J]. Genetics, 2015, 199(2): 413-421.

[45] YOU Q, ZHAI K, YANG D, *et al*.. An E3 ubiquitin ligase-BAG protein module controls plant innate immunity and broad-spectrum disease resistance [J]. Cell Host Microbe, 2016, 20 (6): 758-769.

[46] XING J, NING Q, TANG D, *et al*.. Progress on the role of extrachromosomal DNA in tumor pathogenesis and evolution [J]. Clin. Genet., 2021, 99(4): 503-512.

[47] TORRES G A, GONG Z, IOVENE M, *et al*.. Organization and evolution of subtelomeric satellite repeats in the potato genome [J]. G3 (Bethesda), 2011, 1(2): 85-92.

[48] HE L, LIU J, TORRES G A, *et al*.. Interstitial telomeric re-

peats are enriched in the centromeres of chromosomes in *Solanum* species [J]. Chrom. Res., 2013, 21(1): 5-13.

[49] ALFREDO V, MÉNDEZ-LAGO M, ABAD J P, *et al*.. The birth of the centromere [J]. Cell Cycle (Georgetown, Tex.), 2007, 6(23): 2872-2876.

[50] ZHANG T, TALBERT P B, ZHANG W, *et al*.. The CentO satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres [J]. Proc. Natl. Acad. Sci. USA, 2013, 110(50): 4875-4883.

[51] HAN M, YANG Y, ZHANG M, *et al*.. Considerations regarding centromere assembly in plant whole-genome sequencing [J]. Methods, 2021, 187: 54-56.

[52] HU Y, CHEN J, FANG L, *et al*.. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton [J]. Nat. Genet., 2019, 51 (4): 739-748.

[53] NOVAK P, NEUMANN P, MACAS J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data [J/OL]. BMC Bioinform., 2010, 11: 378[2010-07-15]. https://doi.org/10.1186/1471-2105-11-378.