

植物着丝粒结构及进化的研究进展

刘青

(中国科学院华南植物园, 中国科学院植物资源保护与可持续利用重点实验室, 广州 510650)

摘要: 植物着丝粒是染色体重要结构域, 介导动粒装配。不同物种间着丝粒重复序列快速趋异进化, 着丝粒功能保守, 确保有丝分裂和减数分裂过程中染色体正确分离和准确传递。伴随染色质免疫共沉淀技术(Chromatin immunoprecipitation, ChIP)、ChIP 与高密度芯片相结合技术(ChIP-chip)、ChIP 与高通量测序相结合技术(ChIP-seq)的应用, 植物着丝粒研究获得里程碑式进展: 某些模式植物着丝粒 DNA 序列、蛋白质结构、功能获得大量新认识; 着丝粒基本蛋白质组蛋白 H3 被用来界定着丝粒大小和边界; 某些非着丝粒区域被激活为新着丝粒, 在世代传递中保持稳定性。本文对植物着丝粒结构、功能、进化研究进行了综述, 并探讨了植物着丝粒研究存在的问题。

关键词: 着丝粒; 进化; 特异蛋白质; 趋异进化; 重复序列

doi: 10.11926/j.issn.1005-3395.2015.05.013

Research Progress on Structure and Evolution of Plant Centromeres

LIU Qing

(Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China)

Abstract: The plant centromere is the most important chromosome domain mediating the assembly of kinetochore. The rapid divergent evolution of centromeric repeat sequences and function conservation of centromeres among different species ensure correct segregation and faithful transmission of chromosome in mitosis and meiosis. Along with the development of chromatin immunoprecipitation (ChIP), ChIP-chip, and ChIP-sequencing (ChIP-seq) technologies, three milestone discoveries have achieved in plant centromere research since the last 20 years, such as a lot of new knowledge on the structure, function, and evolution of centromeres from model plants, the fundamental kinetochore protein CENH3 used to delimiting the size and boundaries of centromere, the neocentromeres activated from non-centromeric regions stably transmitted to subsequent generations. The research progress on structure, function, and evolution of plant centromeres are reviewed and the remaining questions of plant centromere studies are discussed.

Keywords: Centromere; Evolution; Specific protein; Divergent evolution; Repeat sequence

植物染色体主缢痕处的特殊分化区域由富含重复序列的异染色质组成, 称为着丝粒。附着在着丝粒两侧各有一个两层蛋白质的盘状结构(动粒外层、内层), 称为动粒(旧称着丝点^[1]), 动粒外层与

纺锤丝微管连接, 动粒内层是动粒与着丝粒核心结构域染色质紧密联系的区域^[2]。如果一段 DNA 序列能够与动粒相互作用, 那么这段序列就是功能着丝粒, 因此所有着丝粒都具有相同的功能——在细

收稿日期: 2014-11-20 接受日期: 2015-04-08

基金项目: 国家自然科学基金项目(31270275); 科技部基础性工作专项(2013FY112100); 中国科学院植物资源保护与可持续利用重点实验室重点项目(201212ZS)资助

作者简介: 刘青(1974~), 女, 副研究员, 主要从事禾本科关键类群系统演化与细胞遗传学研究。E-mail: liuqing@scib.ac.cn

胞分裂期组建动粒^[3]。着丝粒由外到内包括3个结构域：动粒结构域、中央结构域和配对结构域(图1)。由于着丝粒与动粒结构与功能紧密关联，故又称着丝粒/动粒复合体(Centromere/kinetochore complex, CKC)^[4]。动粒结构域位于着丝粒表面，由动粒和围绕动粒外层的纺锤丝微管组成，纺锤丝微管主要由马达蛋白构成，促进染色体运动和分离。中央结构域位于动粒结构域内层，由串联重复序列构成，维持着丝粒/动粒复合体结构的形成。在中央结构域内层，细胞分裂中期两条染色单体在此连接，称为配对结构域。有丝分裂和减数分裂过程中

着丝粒具有两个功能：一个是姐妹染色单体附着位点和分离位点，另一个是控制动粒组装和纺锤丝微管附着，确保染色体正确分离及遗传信息准确传递^[5-6]。伴随染色质免疫共沉淀技术(Chromatin immunoprecipitation, ChIP)、ChIP与高密度芯片相结合技术(ChIP-chip)、ChIP与高通量测序相结合技术(ChIP-seq)的应用，植物着丝粒研究获得里程碑式进展。本文综述植物着丝粒结构、功能和进化研究进展，分析植物着丝粒研究存在的问题，以期对相关研究提供参考。

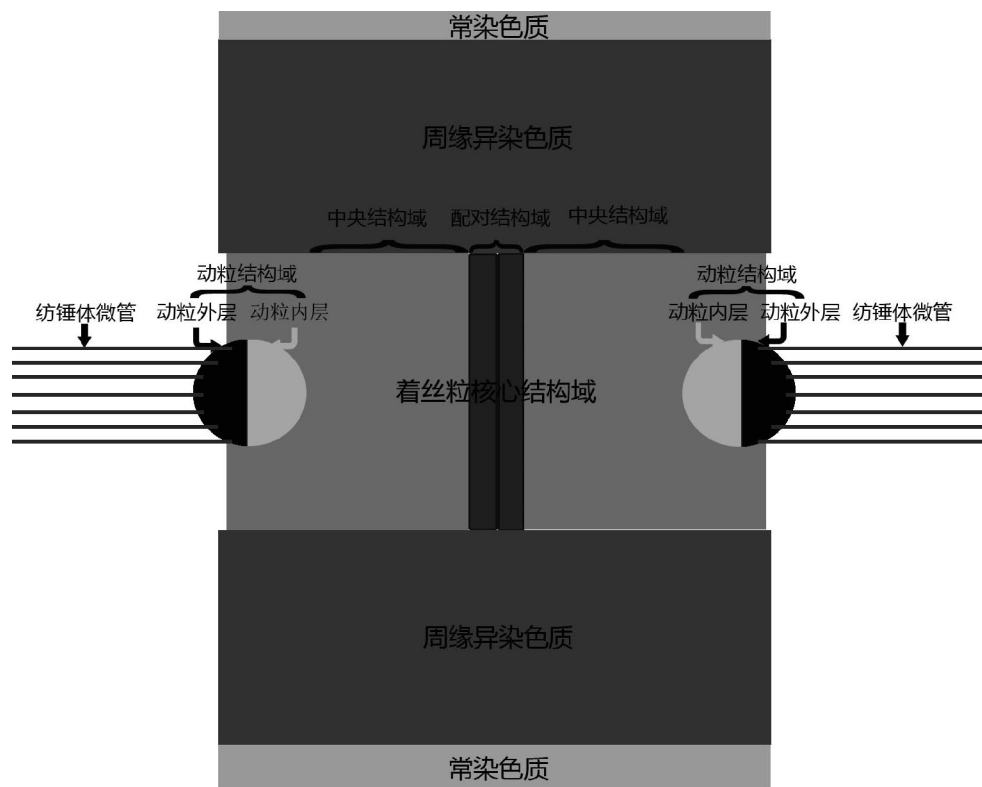


图1 植物中期染色体着丝粒/动粒复合体结构模式图(修改自 Yu 等^[2])

Fig. 1 Linear model of centromere/kinetochore complex of plant metaphase chromosome (Redrawn from Yu et al.^[2])

1 植物着丝粒DNA序列

植物着丝粒DNA序列有3种：(1) 串联重复序列(Tandem repeat, TR)由重复单元构成，重复单元长度稳定，接近单个核小体DNA片段长度(平均约200 bp)或其倍数^[7]，核小体DNA片段长度是着丝粒重复单元长度的倍数^[6,8]；(2) 串联重复序列上散布着拷贝数众多的着丝粒特异的反转录转座子

(Centromeric retrotransposon, CR)，这些序列只特异存在于着丝粒中，在染色体臂上非着丝粒区域鲜有分布^[9-10]；(3) 着丝粒及周缘异染色质区域存在少数具有转录活性的功能基因^[11-12]。着丝粒功能主要在有丝分裂和减数分裂的分裂期完成，而功能基因转录则是在剩余的细胞周期中进行^[13]，因此植物着丝粒重复序列和反转录转座子序列总结见表1，着丝粒区域功能基因见后文。

表 1 已知植物着丝粒重复序列和反转录转座子序列

Table 1 Known plant centromeric repeat and retrotransposon sequences

物种 Species	重复序列 Repeat	大小 Size (bp)	序列类型 Sequence type	染色质免疫共沉淀技术 ChIP	文献 Reference
<i>Arabidopsis arenosa</i>	pAa	166~179	TR	—	[14]
<i>A. gemmifera</i>	pAge1	180	TR	—	[15]
	pAge2	180	TR	—	[16]
<i>A. griffithiana</i>	pAgKB1	180	TR	—	[16]
<i>A. pumila</i>	pApKB2	180	TR	—	[16]
<i>A. thaliana</i>	Repeat family	180	TR	+	[17]
	Athila	10500	RTT	+	[17–18]
<i>Beta corolliflora</i>	pHC8	162	TR	—	[19]
<i>B. procumbens</i>	pTS5	158~160	TR	—	[20]
	pTS4.1	312	TR	—	[20]
	pBp10	417	RTT	—	[19]
<i>B. vulgaris</i>	pBv1	326~327	TR	—	[21]
	pBv26	417	RTT	—	[19]
<i>Brachycome dichromosomatica</i> var. <i>dichromosomatica</i>	Bd49	176	TR	—	[22]
<i>Brachypodium sylvaticum</i>	CCS1	480	RTT	—	[23]
<i>Brassica campestris</i>	pBcKB4	175	TR	—	[24]
<i>B. oleracea</i>	pBoKB1	171	TR	—	[24]
<i>Hordeum vulgare</i>	(AGGGAG) _n satellite	6	TR	+	[25]
	cereba	7176	RTT	+	[25]
<i>Lycopersicon esculentum</i>	TGRIV	7000	RTT	—	[26]
<i>Oryza brachyantha</i>	CentO-F	154	TR	+	[27]
<i>O. rhizomatis</i>	CentO-C1	126	TR	+	[27]
	CentO-C2	366	TR	+	[27]
<i>O. sativa</i>	CentO	155	TR	+	[28]
	CRR	7400~7800	RTT	+	[28–29]
<i>Pennisetum glaucum</i>	pPgKB19	137	TR	—	[30]
<i>Petunia hybrida</i>	pBS-SB1-B5	666	TR	—	[31]
<i>Pinus densiflora</i>	PDCD501	27	TR	—	[32]
<i>Saccharum officinarum</i>	SCEN	140	TR	—	[33]
	CRS	3600	RTT	—	[33]
<i>Secale cereal</i>	Bilby	3400	RTT	—	[34]
<i>Solanum bulbocastanum</i>	pSbTC1	7	TR	—	[35]
<i>S. tuberosum</i>	St49	2754	TR	+	[36]
	St57	1924	TR	+	[36]
	St24	979	TR	+	[36]
	St3-58	2957	TR	—	[36]
	St3-238	3814	TR	—	[36]
	St3-294	5390	TR	—	[36]
	St18	1180	TR	+	[36]
<i>Sorghum bicolor</i>	pSau3A10	137	TR	—	[37]

续表(Continued)

物种 Species	重复序列 Repeat	大小 Size (bp)	序列类型 Sequence type	染色质免疫共沉淀技术 ChIP	文献 Reference
<i>Sorghum bicolor</i>	pSau3A9	745	RTT	—	[37]
<i>Torenia bailonii</i>	BCEN	52	TR	—	[38]
<i>T. fournieri</i>	TCEN	52	TR	—	[38]
<i>Triticum aestivum</i>	Tail	570	TR	—	[39]
	pBS301	250	TR	—	[40]
	CRW	7762~7865	RTT	+	[41]
<i>Vigna unguiculata</i>	pVuKB1	488	TR	—	[42]
<i>Zea mays</i>	CentC	156	TR	+	[43]
	Cent4	740	TR	—	[44]
	CRM	7572	RTT	+	[43]
	B repeat	540	TR	—	[45]
<i>Zingieria biebersteiniana</i>	Zbcen1	593	TR	—	[46]
	Zb47A	263	RTT	—	[46]

TR: 串联重复序列; RTT: 反转录转座子序列; +: 已测试; —: 未测试。
TR: Tandem repeat; RTT: Retrotransposon; +: Detected; —: Not detected.

1.1 重复序列

植物着丝粒串联重复序列(又称卫星 DNA)有 3 个特征:(1) 不同物种的着丝粒串联重复序列长度高度变化,即使同属近缘种间的着丝粒串联重复序列长度也高度变化。拟南芥(*Arabidopsis thaliana*)着丝粒串联重复序列长度为 2.8~4.0 Mb^[47],水稻(*Oryza sativa*)的长度为 65 kb~2 Mb^[28],玉米(*Zea mays*)的长度为 180 kb~2 Mb^[48];(2) 重复序列快速趋异进化。稻属 CC 基因组 3 物种药用野生稻(*O. officinalis*)、*O. eichingeri* 和 *O. rhizomatis*, 着丝粒串联重复序列 CentO (Centromere sequences of *Oryza*)的重复单元分别是 155 bp 的 CentO-C1 和 164 bp 的 CentO-C2, FF 基因组短药野生稻(*O. brachyantha*)重复单元为 154 bp 的 CentO-F,与 CentO-C 两个着丝粒重复单元的序列完全不同,而 CentO-C 两个着丝粒重复单元的 5' 端和 3' 端序列却与相对远缘的玉米着丝粒重复单元序列有显著同源性,暗示禾本科远缘物种间着丝粒串联重复序列可能受到着丝粒功能的选择压力而保持同源性,稻属 CentO-F 趋异进化的机制仍是未解之谜^[18,27];(3) 具有高度有序重复序列(Higher order repeats, HORs)。水稻第 1 号染色体着丝粒(Centromere 1, Cen1)和第 8 号染色体着丝粒(Centromere 8, Cen8)的 HORs 呈现相邻式和间隔式 2 种式样(图 2: A~B)^[49],着丝粒 HORs 呈非随机成簇存在,推测可

能与保留着丝粒特异蛋白质的结合区域及表观遗传修饰位点有关^[50-51]。

1.2 特异反转录转座子序列

着丝粒特异的反转录转座子(CRs)编码反转录酶,以 RNA 为模板自主转录成多拷贝 DNA 序列散布在着丝粒串联重复序列之中,两侧具有长末端重复序列^[37]。例如拟南芥着丝粒重复单元内部含有大量着丝粒特异的反转录转座子 Athila (*Ty3/gypsy*),形成约 3 Mb 的着丝粒中央结构域^[52-54]。植物着丝粒 CRs 序列有 3 个特征:(1) 在科级水平上高度保守。单子叶植物水稻、玉米等禾本科远缘植物中 CRs 保持高达 80% 的序列同源性,是已知最保守的转座元件之一^[37,55];(2) 序列变异具有物种特异性。短花药野生稻(FF 基因组)着丝粒反转录转座子 FRetro3 序列,与水稻的 CRRs、玉米的 CRMs 和大麦的 cereba 等序列完全不同,推测短花药野生稻 CRRs 累积长末端重复序列变异,获得全新的着丝粒反转录转座子序列^[10];(3) 不同区域 CRs 的长末端重复序列不平衡重组概率不同^[9]。不平衡重组频率是指 CRs 单侧长末端重复序列与稳定元件的比率,水稻 *Cen8* 着丝粒区域不平衡重组概率(0.9:1)远低于常染色质区域(2.2:1),而 *Cen8* 中心结构域存在 1 个重组热点区域,在重组热点区域不平衡重组概率较高(2.5:1),表明着丝粒不同区域

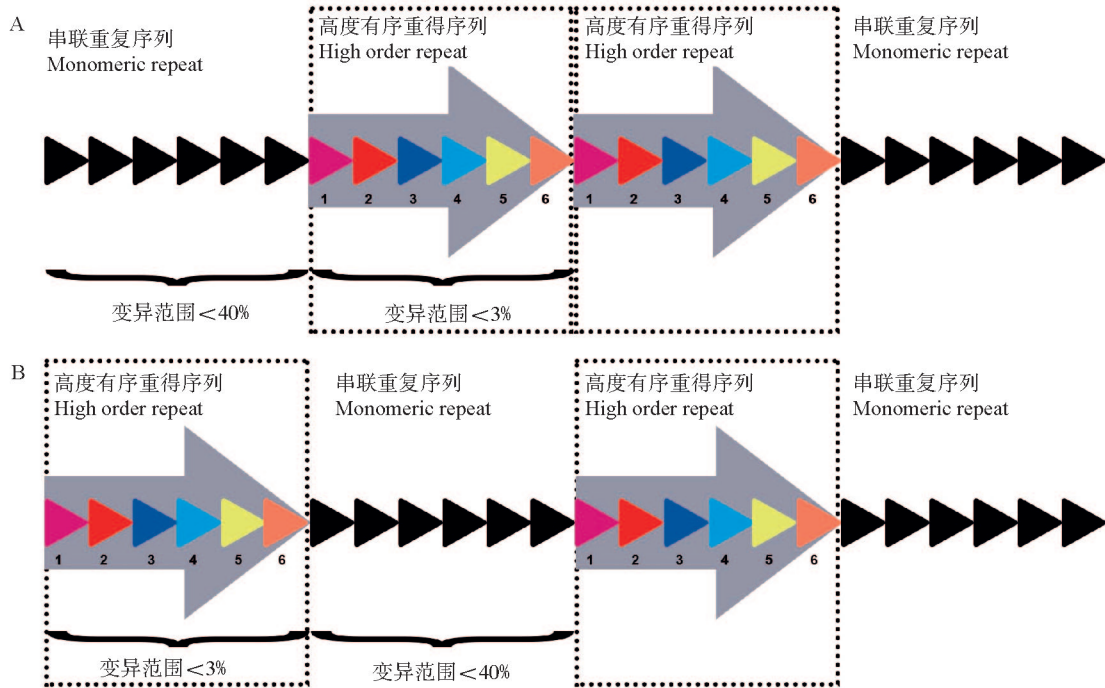


图2 植物着丝粒重复序列结构图。A: 相邻式高度有序重复序列; B: 间隔式高度有序重复序列; ▲: 普通重复单元序列; 1~6: 6个重复单元形成1个高度有序重复序列。(修改自 Lee 等^[49])

Fig. 2 Model of repeat sequences in plant centromeres. A: Adjacent high order repeats (HORs); B: Interval HORs; ▲: Monomeric repeat; 1-6: Each HOR consists of 6 monomeric repeats. (Redrawn from Lee et al.^[49])

CRs 长末端重复序列具有不同变异特性,推测高频率的不平衡重组事件对提高着丝粒重复序列均质化,累积植物特异的着丝粒重复序列有重要意义^[9]。

1.3 功能基因序列

着丝粒及周缘异染色质区域 DNA 序列重组频率较低,在缺少重复序列的着丝粒区域存在少数具有转录活性的功能基因^[11-12]。如拟南芥 *Cen4* 有 5 个活性基因/100 kb、*Cen8* 有 12 个活性基因/100 kb^[50],水稻 *Cen8* 动粒结构域有 4 个活性基因^[13]、短药野生稻 *Cen8* 动粒结构域有 7 个活性基因^[56],马铃薯 (*Solanum tuberosum*) *Cen4*、*Cen6*、*Cen9*、*Cen11* 着丝粒特异组蛋白 CENH3 (Centromere-specific histone H3) 区域有 6 个活性基因^[36]。相对古老的短药野生稻起源时间在中新世中期 10-15 百万年前,与染色体短臂上的功能基因相比,着丝粒功能基因长期承受负(纯化)选择压力,序列分化和突变速率显著降低,暗示维持着丝粒功能基因转录活性的机制,与维持着丝粒重复序列快速趋异进化的机制完全不同,因为着丝粒功能基因在细胞间期具有转录活性,而着丝粒在细胞分裂期发挥功能。研究着丝

粒功能基因何以在长期承受负选择压力后,至今仍具有转录活性是具有挑战性的科学问题^[57]。

2 植物着丝粒蛋白质

植物着丝粒蛋白质(Centromere-specific proteins, CENPs)位于动粒结构域,又称动粒蛋白质。根据植物细胞分裂期着丝粒蛋白质与着丝粒的关系,可分为两种类型(表 2):(1) 瞬时蛋白质(Transient protein)。CENP-E (Centromere protein E)、CENP-F (Centromere protein F)和 CBF5 (Centromere protein F5)等 9 个动粒蛋白质,在有丝分裂中短暂出现,与染色体运动有关;(2) 基本蛋白质(Basic protein)。植物着丝粒特异组蛋白 CENH3、CENP-C (Centromere protein C)等存在于整个细胞周期中,对动粒形成、同源染色体配对、姐妹染色单体复制、联会和分离均有重要作用^[70]。

CENH3 是较早发现的基本蛋白质^[66],是功能着丝粒的基本特征蛋白质,采用染色质免疫共沉淀技术(Chromatin immunoprecipitation, ChIP)鉴别与 CENH3 相互作用的 DNA 序列,确定功能着丝粒的

表 2 已知植物的着丝粒动粒蛋白

Table 2 Known plant kinetochore proteins

类型 Type	植物 Plant	动粒蛋白 Kinetochore component	动粒定位 Kinetochore location	基因克隆 Gene clone	功能 Apparent function	文献 Reference
瞬时蛋白质 Transient protein	大麦, 蚕豆 <i>Hordeum vulgare, Vicia faba</i>	CBF5	是 Yes	是 (大麦) Yes (<i>H. vulgare</i>)	未知 Unknown	[58]
	大麦, 蚕豆 <i>H. vulgare, V. faba</i>	CENP-E	是 Yes	无 No	染色体运动 Chromosome motility	[59]
	大麦 <i>H. vulgare</i>	CENP-F	是 Yes	无 No	未知 Unknown	[58,60]
	玉米 <i>Zea mays</i>	MAD2	是 Yes	是 Yes	纺锤体关卡蛋白 Spindle checkpoint	[61]
	蚕豆 <i>V. faba</i>	MPM2	是 Yes	无 No	未知 Unknown	[62]
	拟南芥 <i>Arabidopsis thaliana</i>	ZW10	无 No	是 Yes	纺锤体关卡蛋白 Spindle checkpoint	[63]
	玉米 <i>Z. mays</i>	3F3/2 antigen	是 Yes	无 No	纺锤体关卡蛋白 Spindle checkpoint	[61]
	大蒜, 紫娇花 <i>Allium sativum, Tulbaghia violacea</i>	6C6 antigen	是 Yes	无 No	微管组织中心蛋白 MTOC	[64]
	蚕豆 <i>V. faba</i>	γ -Tubulin	是 Yes	无 No	微管组织中心蛋白 MTOC	[65]
	基本蛋白质 Basic protein	拟南芥, 烟草, 玉米 <i>A. thaliana, Nicotiana tabacum, Z. mays</i>	CENH3	是 Yes	是 Yes	结构蛋白质 Constitutive
大麦, 蚕豆, 玉米 <i>H. vulgare, V. faba, Z. mays</i>		CENP-C	是 Yes	是 (玉米) Yes (<i>Z. mays</i>)	结构蛋白质 Constitutive	[58,67]
麝香百合 <i>Lilium longiflorum</i>		Meiotic histone	是 Yes	是 Yes	未知 Unknown	[68]
拟南芥, 大麦, 蚕豆 <i>A. thaliana, H. vulgare, V. faba</i>		SKP1 (CENP-C homologs)	是 Yes	是 Yes	未知 Unknown	[58,69]

边界和大小^[71]。CENH3 在种级和亚种级水平快速进化,其氨基端尾部和蛋白质折叠域皆可变异,拟南芥属(*Arabidopsis*)植物的 CENH3 氨基端氨基酸序列组成及长度高度可变,推测与其他着丝粒蛋白质相互作用密切相关^[72], CENH3 羧基端蛋白质折叠域的变异与核小体定位密切相关^[73-74]。CENH3 具有 3 个方面的应用:(1) 确定功能着丝粒边界和大小。水稻第 8 号染色体着丝粒中与 CENH3 结合的卫星 DNA (CentO)约 750 kb,则功能着丝粒大小约为 750 kb^[13];(2) 鉴定着丝粒 DNA 序列。拟南芥的 ChIP 研究表明卫星 DNA 和 CRs 能被 CENH3 抗体沉淀,说明卫星 DNA 和 CRs 都能与 CENH3 发生相互作用,是拟南芥着丝粒的功能组分^[17];(3) 介导单亲基因组消除(*Uniparental genome elimination*)来获得单倍个体,在大麦(*Hordeum vulgare*, 母本)× 球茎大麦(*H. bulbosum*, 父本)杂种胚中,球茎大麦 CENH3 蛋白质丢失,导致来自球茎大麦染色体的着丝点在有丝分裂过程中无活性,从

而形成单倍体^[75],该方法将基因工程技术与常规杂交育种手段相结合,达到单倍体诱导的目的^[76]。

CENP-C 对动粒结构的正确组装和中期、后期转化是必须的,玉米、拟南芥、高粱(*Sorghum bicolor*)、甘蔗(*Saccharum officinarum*)等植物着丝粒蛋白质 CENP-C 都具有 1 个保守的 24 氨基酸残基组成的基序^[77],代表 CENP-C 与 CENH3 交互作用的保守区域^[78]。CENP-C 氨基(5')端第 1~6 外显子、羧基(3')端第 13~14 外显子区段和氨基酸编码序列高度保守,而在第 9~10 和第 11~12 外显子区段具有 2 种倍增模式,倍增区段的高度同源序列暗示倍增事件起源相对古老。动物、植物 CENP-C 长度相似,保守区域排列顺序一致。玉米和甘蔗 CENP-C 经历正选择(快速进化),拟南芥 CENH3 也经历正选择,尤其是与着丝粒串联重复序列绑定的区域,推测动粒基本蛋白质与着丝粒重复序列均能以快速进化方式而紧密结合,并维持染色体的正常功能,是否存在同一种驱动力导致 CENP-C

氨基端基序以外区域和 CENH3 快速进化,尚不知晓^[66,72,79]。

植物全部着丝粒蛋白质的组成及分子机制尚不完全清楚,值得深入研究的问题有:CENPs 组装机制和表达调控过程、CENPs 与着丝粒重复序列结合过程、CENPs 在细胞分裂过程中的动力学特征等。进一步解析植物着丝粒蛋白质结构与功能,对于认识着丝粒染色质特征及其与着丝粒功能的关系将会有极大帮助。

3 植物着丝粒进化机制

大多数植物着丝粒和近着丝粒区域都包含重复序列,对于是何种机制维持着丝粒重复序列均质化,目前有 4 种假说。第一种是库假说,推测着丝粒有数套不同丰度的重复单元,伴随突变不断产生新的重复单元,遗传漂变累积新产生的重复单元,借助不平衡重组事件实现着丝粒重复序列同质化^[80]。第二种是减数分裂驱动假说,推测着丝粒重复序列和与之结合的着丝粒动粒蛋白质之间存在亲和力,亲和力的选择驱动促使着丝粒重复序列均质化,即某些着丝粒重复序列具有与动粒蛋白质优先结合的亲和力,被优先分离到功能大孢子中,产生与动粒蛋白质结合力越来越强的着丝粒重复序列而被快速固定,反之动粒蛋白质快速进化降而低与之结合的着丝粒重复序列的亲和力,削弱着丝粒重复序列对动粒大小的影响,促使具有一致结构的 HORs 出现^[66,81]。第三种是致同进化假说。Melters 等^[6]基于 282 种动物、植物着丝粒串联重复序列的系统发育研究表明,植物着丝粒串联重复序列快速驱异进化表现在重复单元长度、GC 含量和丰度皆可变,禾本科着丝粒串联重复序列呈现相对一致的保守性和快速进化模式,玉米着丝粒串联重复序列与黍属(*Panicum*)、狗尾草属(*Setaria*)、甚至稻属(41 百万年前分化)相对同源,形成鲜明对比的是近缘属高粱属(*Sorghum*)-玉米属(9 百万年前分化)、大麦属-山羊草属(*Aegilops*)(14 百万年前分化)之间几乎没有同源性,推测是重复单元承受的功能压力导致着丝粒重复序列致同进化。第四种是突然进化假说。利用 ChIP-seq 技术对马铃薯着丝粒序列进行研究,12 条染色体上着丝粒 DNA 序列组成完全不同,其中 5 条染色体着丝粒区域 DNA 由不同的单拷贝或低拷贝 DNA 序列组成,另外 6 条染色体

上由不同的高度重复序列组成,这些重复序列在马铃薯近缘野生种均不存在,因此推测着丝粒 DNA 序列从非重复序列进化到重复序列可能是以突然的方式完成,而不是在新着丝粒区域逐步积累重复单元形成的^[36],该假说在植物中为首次报道。

某些非着丝粒区域在一定条件下获得着丝粒活性而产生新着丝粒^[82-83],如大麦第 7 号染色体短臂^[82]、玉米第 3 号染色体短臂^[83]。新着丝粒与正常着丝粒位置较近,缺乏物种特异的着丝粒重复序列及反转录转座子序列,在世代传递中具有稳定性。新着丝粒形成假说有 2 种:第一种是着丝粒断裂假说(Centromere breakage)。对大麦第 7 号染色体着丝粒动粒装配的研究表明,这种装配过程仅仅涉及一部分着丝粒重复序列,着丝粒特异蛋白质向周缘异染色质区域迁移,一旦着丝粒断裂,两端染色体臂各自携带一部分着丝粒重复序列,形成 2 条端着丝粒染色体,有丝分裂和减数分裂中功能正常^[25,84]。第二种是着丝粒重新定位假说(Centromere repositioning event)。对黄瓜(*Cucumis sativus*)第 7 号染色体及甜瓜(*C. melo*)第 2 号染色体的研究表明,着丝粒激活或者失活过程中,均伴随着丝粒周缘异染色质获得或者丢失,着丝粒失活后丢失周缘异染色质部分序列,而新着丝粒激活后获得周缘异染色质部分序列,推测周缘异染色质序列对着丝粒重新定位起关键作用^[85]。新着丝粒激活在动物中往往是致命的^[86],暗示新着丝粒对于动物似乎毫无进化优势可言。如果植物新着丝粒没有进化优势,在居群中将很快被淘汰,而不会出现在世代传递中保持稳定性的新着丝粒染色体,如大麦第 7 号染色体短臂、玉米第 3 号染色体短臂等。有研究表明着丝粒失活、新着丝粒激活将影响所在着丝粒区域、着丝粒周缘异染色质区域的基因重组^[87],毫无疑问新着丝粒会导致染色体形态、核型发生巨大变化,其进化优势是否在于调控所在区域的基因表达变化尚需深入研究^[57,88]。

4 植物着丝粒研究存在的问题

1990 年以来的 20 多年间,伴随染色质免疫共沉淀技术(ChIP)、高密度芯片(ChIP-chip)和高通量测序(ChIP-seq)技术的发展,植物着丝粒研究获得里程碑式的进展:第一是模式植物(单子叶植物水稻、玉米、高粱等,双子叶植物拟南芥、马铃薯、

黄瓜等)着丝粒 DNA 序列、蛋白质结构、功能获得大量新认识^[13,36-37,67,69,85];第二是着丝粒基本蛋白质 CENH3 用来界定着丝粒 / 动粒复合体的大小和边界^[13];第三是某些非着丝粒区域被激活为新着丝粒,在世代传递中保持稳定^[89]。虽然目前已得到部分植物的着丝粒序列,但对它们的认识和分析仍停留在起步阶段,各种着丝粒功能模型层出不穷,归纳起来植物着丝粒研究存在的问题有:(1) 除极少数物种外,低等植物着丝粒研究鲜见报道^[6];(2) 颈卵器植物(如苔藓、蕨类)着丝粒研究鲜见报道;(3) 高等植物非典型着丝粒研究少见报道,目前仅报道有马铃薯着丝粒重复单元长 979~5390 bp^[36],与植物着丝粒常见的重复单元(150~180 bp)大小显著不同;(4) 着丝粒序列和功能之间的进化关系缺乏系统研究,目前已证实一方面植物着丝粒序列即使在亲缘关系很近的物种之间高度多样,另一方面着丝粒功能高度保守,但着丝粒序列本身不编码蛋白质,缺乏进化依存关系,迄今着丝粒结构和功能之间的进化关系悬而未决;(5) 新着丝粒的进化机制还不清楚,新着丝粒与正常着丝粒的共性特征包括:形成细胞学上可辨认的收缩、结合着丝粒特异蛋白质、在整个细胞循环中能够稳定传递,而与正常着丝粒缺乏共同序列,新着丝粒的出现暗示着丝粒功能可能被独立于 DNA 序列的外源因素调节,系统研究失活的着丝粒区域以及激活后的新着丝粒区域的结构和功能,将科学地解释新着丝粒产生的进化优势;(6) 着丝粒功能模型需要提出新思想,如着丝粒功能可能依赖重复单元;着丝粒 DNA 序列和蛋白质之间存在共进化机制^[90],即着丝粒不需要特定重复单元而达到自我复制。多种植物着丝粒的研究资料提出了有价值的信息,着丝粒功能模型在重复单元依赖模型与完全不需要的极端模型之间可能需要一个新的综合。伴随新技术和物种附加系分析方法的开发应用,必将深化对着丝粒序列均质化和着丝粒进化机制的认识。

致谢 感谢福建农林大学王凯教授就着丝粒 DNA 序列提供宝贵意见。

参考文献

- [1] Hong X J. Centromeres and their evolution [J]. J SW China Norm Univ, 1988(1): 108-114.
洪锡钧. 着丝点及其进化 [J]. 西南师范大学学报, 1988(1): 108-114.
- [2] Yu H G, Hiatt E N, Dawe R K. The plant kinetochore [J]. Trends Plant Sci, 2000, 5(12): 543-547.
- [3] Cleveland D W, Mao Y H, Sullivan K F. Centromeres and kinetochores: From epigenetics to mitotic checkpoint signaling [J]. Cell, 2003, 112(4): 407-421.
- [4] Brinkley B R, Valdivia M M, Tousson A, et al. The kinetochore: Structure and molecular characterization [C]// Hyams J, Brinkley B R. Mitosis: Moleculars and Mechanisms. New York: Academic Press, 1989: 77-118.
- [5] Wang G X, Zhang X W, Jin W W. An overview of plant centromeres [J]. J Genet Genom, 2009, 36(9): 529-537.
王桂香, 张学勇, 金危危. 植物着丝粒研究进展 [J]. 遗传学报, 2009, 36(9): 529-537.
- [6] 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]. Genome Biol, 2013, 14(1): R10.
- [7] Wu J Z, 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.
- [8] Henikoff S, Ahmad K, Malik H S. The centromere paradox: Stable inheritance with rapidly evolving DNA [J]. Science, 2001, 293(5532): 1098-1102.
- [9] Ma J X, Bennetzen J L. Recombination, rearrangement, reshuffling, and divergence in a centromere region of rice [J]. Proc Natl Acad Sci USA, 2006, 103(2): 383-388.
- [10] Gao D Y, Gill N, Kim H R, et al. A lineage-specific centromere retrotransposon in *Oryza brachyantha* [J]. Plant J, 2009, 60(5): 820-831.
- [11] Yan H H, Talbert P B, Lee H R, et al. Intergenic locations of rice centromeric chromatin [J]. PLoS Biol, 2008, 6(11): e286.
- [12] Wu Y F, Kikuchi S, Yan H H, et al. Euchromatic subdomains in rice centromeres are associated with genes and transcription [J]. Plant Cell, 2011, 23(11): 4054-4064.
- [13] Nagaki K, Cheng Z K, Ouyang S, et al. Sequencing of a rice centromere uncovers active genes [J]. Nat Genet, 2004, 36(2): 138-145.
- [14] Kamm A, Galasso I, Schmidt T, et al. Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationships between *Arabidopsis* species [J]. Plant Mol Biol, 1995, 27(5): 853-862.
- [15] Kawabe A, Nasuda S. Structure and genomic organization of centromeric repeats in *Arabidopsis* species [J]. Mol Gen Genom, 2005, 272(6): 593-602.
- [16] Heslop-Harrison J S, Brandes A, Schwarzacher T. Tandemly repeated DNA sequences and centromeric chromosomal regions of *Arabidopsis* species [J]. Chromosome Res, 2003, 11(3): 241-253.
- [17] Nagaki K, Song J Q, Stupar R M, et al. Molecular and cytological

- analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres [J]. *Genetics*, 2003, 163(2): 759–770.
- [18] Péliissier T, Tutois S, Deragon J M, et al. Athila, a new retroelement from *Arabidopsis thaliana* [J]. *Plant Mol Biol*, 1995, 29(3): 441–452.
- [19] Gindullis F, Desel C, Galasso I, et al. The large-scale organization of the centromeric region in *Beta* species [J]. *Genome Res*, 2001, 11(2): 253–265.
- [20] Schmidt T, Heslop-Harrison J S. High-resolution mapping of repetitive DNA by *in situ* hybridization: Molecular and chromosomal features of prominent dispersed and discretely localized DNA families from the wild beet species *Beta procumbens* [J]. *Plant Mol Biol*, 1996, 30(6): 1099–1113.
- [21] Schmidt T, Metzclaff M. Cloning and characterization of a *Beta vulgaris* satellite DNA family [J]. *Gene*, 1991, 101(2): 247–250.
- [22] Leach C R, Donald T M, Franks T K, et al. Organisation and origin of a B chromosome centromeric sequence from *Brachycome dichromosomatica* [J]. *Chromosoma*, 1995, 103(10): 708–714.
- [23] Aragón-Alcaide L, Miller T, Schwarzacher T, et al. A cereal centromeric sequence [J]. *Chromosoma*, 1996, 105(5): 261–268.
- [24] Harrison G E, Heslop-Harrison J S. Centromeric repetitive DNA sequences in the genus *Brassica* [J]. *Theor Appl Genet*, 1995, 90(2): 157–165.
- [25] 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 [J]. *Chromosoma*, 2007, 116(3): 275–283.
- [26] Chang S B, Yang T J, Datema E, et al. FISH mapping and molecular organization of the major repetitive sequences of tomato [J]. *Chromosome Res*, 2008, 16(7): 919–933.
- [27] Lee H R, Zhang W L, 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.
- [28] Cheng Z K, Dong F G, 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.
- [29] Dong F G, Miller J T, Jackson S A, et al. Rice (*Oryza sativa*) centromeric regions consist of complex DNA [J]. *Proc Natl Acad Sci USA*, 1998, 95(14): 8135–8140.
- [30] Kamm A, Schmidt T, Heslop-Harrison J S. Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum* [J]. *Mol Gen Genet*, 1994, 244(4): 420–425.
- [31] Entani T, Iwano M, Shiba H, et al. Centromeric localization of an S-RNase gene in *Petunia hybrida* Vilm. [J]. *Theor Appl Genet*, 1999, 99(3/4): 391–397.
- [32] Hizume M, Shibata F, Maruyama Y, et al. Cloning of DNA sequences localized on proximal fluorescent chromosome bands by microdissection in *Pinus densiflora* Sieb. & Zucc. [J]. *Chromosoma*, 2001, 110(5): 345–351.
- [33] Nagaki K, Murata M. Characterization of CENH3 and centromere-associated DNA sequences in sugarcane [J]. *Chromosome Res*, 2005, 13(2): 195–203.
- [34] Francki M G. Identification of *Bilby*, a diverged centromeric Ty1-copia retrotransposon family from cereal rye (*Secale cereale* L.) [J]. *Genome*, 2001, 44(2): 266–274.
- [35] Tek A L, Jiang J M. The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence [J]. *Chromosoma*, 2004, 113(2): 77–83.
- [36] Gong Z Y, Wu Y F, Koblížková A, et al. Repeatless and repeat-based centromeres in potato: Implications for centromere evolution [J]. *Plant Cell*, 2012, 24(9): 3559–3574.
- [37] Miller J T, Jackson S A, Nasuda S, et al. Cloning and characterization of a centromere-specific repetitive DNA element from *Sorghum bicolor* [J]. *Theor Appl Genet*, 1998, 96(6–7): 832–839.
- [38] Kikuchi S, Kishii M, Shimizu M, et al. Centromere-specific repetitive sequences from *Torenia*, a model plant for interspecific fertilization, and whole-mount FISH of its interspecific hybrid embryos [J]. *Cytogenet Genome Res*, 2005, 109(1/2/3): 228–235.
- [39] Kishii M, Nagaki K, Tsujimoto H. A tandem repetitive sequence located in the centromeric region of common wheat (*Triticum aestivum*) chromosomes [J]. *Chromosome Res*, 2001, 9(5): 417–428.
- [40] Cheng Z J, Murata M. A centromeric tandem repeat family originating from a part of *Ty3/gypsy*: Retroelement in wheat and its relatives [J]. *Genetics*, 2003, 164(2): 665–672.
- [41] Liu Z, Yue W, Li D Y, et al. Structure and dynamics of retrotransposons at wheat centromeres and pericentromeres [J]. *Chromosoma*, 2008, 117(5): 445–456.
- [42] Goel S, Raina S N, Ogihara Y. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of nuclear ribosomal DNA in the *Phaseolus-Vigna* complex [J]. *Mol Phylogenet Evol*, 2002, 22(1): 1–19.
- [43] 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.
- [44] Page B T, Wanous M K, Birchler J A. Characterization of a maize chromosome 4 centromeric sequence: Evidence for an evolutionary relationship with the B chromosome centromere [J]. *Genetics*, 2001, 159(1): 291–302.
- [45] Alfenito M R, Birchler J A. Molecular characterization of a maize B chromosome centric sequence [J]. *Genetics*, 1993, 135(2): 589–597.

- [46] Saunders V A, Houben A. The pericentromeric heterochromatin of the grass *Zingera biebersteiniana* ($2n=4$) is composed of Zbcen1-type tandem repeats that are intermingled with accumulated dispersedly organized sequences [J]. *Genome*, 2001, 44(6): 955–961.
- [47] Martinez-Zapater J M, Estelle M A, Somerville C R. A high repeated sequence in *Arabidopsis thaliana* [J]. *Mol Gen Genet*, 1986, 204(3): 417–423.
- [48] Jin W W, Melo J R, Nagaki K, et al. Maize centromeres: Organization and functional adaptation in the genetic background of oat [J]. *Plant Cell*, 2004, 16(3): 571–581.
- [49] Lee H R, Neumann P, Macas J, et al. Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice [J]. *Mol Biol Evol*, 2006, 23(12): 2505–2520.
- [50] Copenhaver G P, Nickel K, Kuromori T, et al. Genetic definition and sequence of *Arabidopsis* centromeres [J]. *Science*, 1999, 286(5449): 2468–2474.
- [51] Houben A, Schubert I. DNA and proteins of plant centromeres [J]. *Curr Opin Plant Biol*, 2003, 6(6): 554–560.
- [52] Langdon T, Seago C, Mende M, et al. Retrotransposon evolution in diverse plant genomes [J]. *Genetics*, 2000, 156(1): 313–325.
- [53] Hosouchi T, Kumekawa N, Tsuruoka H, et al. Physical map-based sizes of the centromeric regions of *Arabidopsis thaliana* chromosomes 1, 2, and 3 [J]. *DNA Res*, 2002, 9(4): 117–121.
- [54] Lim K B, Yang T J, Hwang Y J, et al. Characterization of the centromere and peri-centromere retrotransposons in *Brassica rapa* and their distribution in related *Brassica* species [J]. *Plant J*, 2007, 49(2): 173–183.
- [55] Jackson S A, Jiang J, Friebe B, et al. Structure of the rye midget chromosome analyzed by FISH and C-banding [J]. *Genome*, 1997, 40(5): 782–784.
- [56] Fan C Z, Walling J G, Zhang J W, et al. Conservation and purifying selection of transcribed genes located in a rice centromere [J]. *Plant Cell*, 2011, 23(8): 2821–2830.
- [57] Jiang J M. Rice centromeres [C]// Jiang J M, Birchler J A. *Plant Centromere Biology*. Iowa: John Wiley & Sons, Inc., 2013: 15–24.
- [58] ten Hoopen R, Manteuffel R, Doležel J, et al. Evolutionary conservation of kinetochore protein sequences in plants [J]. *Chromosoma*, 2000, 109(7): 482–489.
- [59] ten Hoopen R, Schleker T, Manteuffel R, et al. Transient CENP-E-like kinetochore proteins in plants [J]. *Chromosome Res*, 2002, 10(7): 561–570.
- [60] Rattner J B, Rao A, Fritzler M J, et al. CENP-F is a ca. 400 kDa kinetochore protein that exhibits a cell-cycle dependent localization [J]. *Cell Motil Cytoskeleton*, 1993, 26(3): 214–226.
- [61] Yu H G, Muszynski M G, Dawe R K. The maize homologue of the cell cycle checkpoint protein MAD2 reveals kinetochore substructure and contrasting mitotic and meiotic localization patterns [J]. *J Cell Biol*, 1999, 145(3): 425–435.
- [62] Binarova P, Cihalikova J, Dolezel J. Localization of MPM-2 recognized phosphoproteins and tubulin during cell cycle progression in synchronized *Vicia faba* root meristem cells [J]. *Cell Biol Int*, 1993, 17(9): 847–856.
- [63] Starr D A, Williams B C, Li Z X, et al. Conservation of the centromere/kinetochore protein ZW10 [J]. *J Cell Biol*, 1997, 138(6): 1289–1301.
- [64] Schmit A C, Stoppin V, Chevrier V, et al. Cell cycle dependent distribution of a centrosomal antigen at the perinuclear MTOC or at the kinetochores of higher plant cells [J]. *Chromosoma*, 1994, 103(5): 343–351.
- [65] Binarová P, Hause B, Doležel J, et al. Association of γ -tubulin with kinetochore/centromeric region of plant chromosomes [J]. *Plant J*, 1998, 14(6): 751–757.
- [66] 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.
- [67] Dawe R K, Reed L M, Yu H G, et al. A maize homolog of mammalian CENPC is a constitutive component of the inner kinetochore [J]. *Plant Cell*, 1999, 11(7): 1227–1238.
- [68] Suzuki T, Ide N, Tanaka I. Immunocytochemical visualization of the centromeres during male and female meiosis in *Lilium longiflorum* [J]. *Chromosoma*, 1997, 106(7): 435–445.
- [69] Yang M, Hu Y, Lodhi M, et al. The *Arabidopsis* *SKP1-LIKE1* gene is essential for male meiosis and may control homologue separation [J]. *Proc Natl Acad Sci USA*, 1999, 96(20): 11416–11421.
- [70] Chan G K, Liu S T, Yen T J. Kinetochore structure and function [J]. *Trends Cell Biol*, 2005, 15(11): 589–598.
- [71] Nagaki K, Terada K, Wakimoto M, et al. Centromere targeting of alien CENH3s in *Arabidopsis* and tobacco cells [J]. *Chromosome Res*, 2010, 18(2): 203–211.
- [72] Cooper J L, Henikoff S. Adaptive evolution of the histone fold domain in centromeric histones [J]. *Mol Biol Evol*, 2004, 21(9): 1712–1718.
- [73] Ravi M, Chan S W. Haploid plants produced by centromere-mediated genome elimination [J]. *Nature*, 2010, 464(7288): 615–618.
- [74] Lermontova I, Rutten T, Schubert I. Deposition, turnover, and release of CENH3 at *Arabidopsis* centromeres [J]. *Chromosoma*, 2011, 120(6): 633–640.
- [75] Sanei M, Pickering R, Kumke K, et al. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids [J]. *Proc Natl Acad Sci USA*, 2011, 108(33): E498–E505.
- [76] Lermontova I, Schubert I. CENH3 for establishing and maintaining

- centromeres [C]// Jiang J M, Birchler J A. Plant Centromere Biology. Iowa: John Wiley & Sons, Inc., 2013: 67–82.
- [77] Talbert P B, Bryson T D, Henikoff S. Adaptive evolution of centromere proteins in plants and animals [J]. *J Biol*, 2004, 3(4): 18.
- [78] Luger K, Mader A W, Richmond R K, et al. Crystal structure of the nucleosome core particle at 2.8 Å resolution [J]. *Nature*, 1997, 389(6648): 251–260.
- [79] She C W, Song Y C. Advances in research of the structure and function of plant centromeres [J]. *Hereditas*, 2006, 28(12): 1597–1606.
- 余朝文, 宋运淳. 植物着丝粒结构和功能的研究进展 [J]. *遗传*, 2006, 28(12): 1597–1606.
- [80] Ugarković D, Plohl M. Variation in satellite DNA profiles: Causes and effects [J]. *EMBO J*, 2002, 21(22): 5955–5959.
- [81] Yi C D, Gu M H. Advances in research of plant centromere [J]. *J Yangzhou Univ (Agri Life Sci)*, 2007, 28(1): 39–44.
- 裔传灯, 顾铭洪. 植物着丝粒的研究进展 [J]. *扬州大学学报: 农业与生命科学版*, 2007, 28(1): 39–44.
- [82] Voullaire L E, Slater H R, Petrovic V, et al. A functional marker centromere with no detectable alphasatellite, satellite III, or CENP-B protein: activation of a latent centromere [J]. *Amer J Hum Genet*, 1993, 52(6): 1153–1163.
- [83] Topp C N, Okagaki R J, Melo J R, et al. Identification of a maize neocentromere in an oat-maize addition line [J]. *Cytogenet Genome Res*, 2009, 124(3–4): 228–238.
- [84] Nasuda S, Hudakova S, Schubert I, et al. Stable barley chromosomes without centromeric repeats [J]. *Proc Nat Acad Sci USA*, 2005, 102(28): 9842–9847.
- [85] Han Y H, Zhang Z H, Liu C X, et al. Centromere repositioning in *Cucurbit* species: Implication of the genomic impact from centromere activation and inactivation [J]. *Proc Natl Acad Sci USA*, 2009, 106(35): 14937–14941.
- [86] Marshall O J, Chueh A C, Wong L H, et al. Neocentromeres: New insights into centromere structure, disease development, and karyotype evolution [J]. *Amer J Hum Genet*, 2008, 82(2): 261–282.
- [87] Zhang W L, Friebe B, Gill B S, et al. Centromere inactivation and epigenetic modifications of a plant chromosome with three functional centromeres [J]. *Chromosoma*, 2010, 119(5): 553–563.
- [88] Ding G, Yao N, Wu Q, et al. Research progress on the structure and function of centromeres [J]. *Chin Bull Bot*, 2008, 25(2): 149–160.
- 丁戈, 姚南, 吴琼, 等. 着丝粒结构与功能研究的新进展 [J]. *植物学通报*, 2008, 25(2): 149–160.
- [89] Jiang J M. Centromere evolution [C]// Jiang J M, Birchler J A. Plant Centromere Biology. Iowa: John Wiley & Sons, Inc., 2013: 159–169.
- [90] Hall S E, Kettler G, Preuss D. Centromere satellites from *Arabidopsis* populations maintenance of conserved and variable domains [J]. *Genome Res*, 2003, 13(2): 195–205.