Plant Centromere Biology

Edited by Jiming Jiang and James A. Birchler

WILEY-BLACKWELL

Table of Contents

Plant Centromere Biology

Editors
JIMING JIANG
JAMES A. BIRCHLER

WILEY-BLACKWELL
A John Wiley & Sons, Inc., Publication

This edition first published 2013 © 2013 by John Wiley & Sons, Inc.

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Editorial offices: 2121 State Avenue, Ames, Iowa 50014-8300, USA

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

Plant centromere biology / editors, Jiming Jiang, James A. Birchler.

pages cm

Includes bibliographical references and index.

ISBN 978-1-119-94921-3 (hardback : alk. paper) 1. Plant molecular genetics. 2. Plant genomes. 3. Plant physiology. 4. Centromere. 5. Botany. I. Jiang, Jiming, editor of compilation. II. Birchler, James A. (James Arthur), 1950- editor of compilation.

QK981.4P53 2013

572.8'2–dc23

2012038414

A catalogue record for this book is available from the British Library.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Cover images are provided by James Birchler, Zhiyun Gong, and Minoru Murata.

Cover design by Modern Alchemy LLC

Contributors

James A. Birchler Division of Biological Sciences, University of Missouri, Columbia, Missouri, USA

Simon W.-L. Chan[*] Department of Plant Biology, University of California, Davis, California, USA

Ryan N. Douglas Division of Biological Sciences, University of Missouri, Columbia, Missouri, USA

Lila Fishman Division of Biological Sciences, University of Montana, Missoula, Montana, USA

Bernd Friebe Wheat Genetic and Genomic Resources Center and Department of Plant Pathology, Throckmorton Plant Sciences Center, Kansas State University, Manhattan, Kansas, USA

Robert T. Gaeta Division of Biological Sciences, University of Missouri, Columbia, Missouri, USA

Bikram S. Gill Wheat Genetic and Genomic Resources Center and Department of Plant Pathology, Throckmorton Plant Sciences Center, Kansas State University, Manhattan, Kansas, USA

Stefan Heckmann Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Andreas Houben Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Jiming Jiang Department of Horticulture, University of Wisconsin, Madison, Wisconsin, USA

Inna Lermontova Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Adam J. Lukaszewski Department of Botany and Plant Science, University of California, Riverside, California, USA

Minoru Murata Research Institute of Plant Science and Resources, Okayama University, Kurashiki, Japan

Gernot Presting Molecular Biosciences and Bioengineering, University of Hawaii, Honolulu, Hawaii, USA

María J. Puertas Departamento de Genética, Facultad de Biología, Universidad Complutense, Madrid, Spain

Lili Qi USDA-ARS, Northern Crop Science Laboratory, Fargo, North Dakota, USA

Maruthachalam Ravi Department of Plant Biology, University of California, Davis, California, USA

Thomas Schmidt Department of Biology, Dresden University of Technology, Dresden, Germany

Ingo Schubert Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Alfredo Villasante Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Universidad Autónoma de Madrid, Madrid, Spain

Beatrice Weber Department of Biology, Dresden University of Technology, Dresden, Germany

Falk Zakrzewski Department of Biology, Dresden University of Technology. Dresden, Germany

* Deceased.

Preface

The centromere is the chromosomal domain that directs the formation of the kinetochore, a proteinaceous structure that interacts with the spindle microtubules to ensure proper chromosomal segregation. The centromere appears as a “primary constriction” on the metaphase chromosomes and can be readily distinguished from the rest of the chromosome. Centromeres were described more than 100 years ago. Yet there was very little information available about the DNA and proteins associated with centromeres in higher eukaryotes before 1990, due to the incredibly complex structure of this unique chromosomal domain. However, remarkable progress was achieved in centromere research in the last 20 years. There were several milestone discoveries: (a) centromeres contain a unique histone H3 variant CENP-A (CID in *Drosophila*, CENH3 in plants), which is the functional mark of centromeres; (a) neocentromere formation: new centromeres can be activated from non-centromeric DNA by recruiting the CENP-A to the new location; and (c) developing artificial chromosomes using cloned centromeric DNA. All of these milestone discoveries were made in model animal species.

Several classical discoveries of centromere function were made in plants. Marcus Rhoades reported the first “neocentromere function” of a heterochromatic knob of a maize chromosome in 1942. C.D. Darlington discovered centromere misdivision in 1939, and in the early 1950s Ernest Sears discovered that both parts of the divided centromeres of wheat chromosomes retain function; thus, a centromere must consist of several units that are equally functional. Several plant species have been established as unique models in centromere research. A number of novel

discoveries on the structure, function, and evolution of centromeres have been made using these plant models. For example, centromeres in most higher eukaryotes contain exclusively long arrays of satellite repeats. However, several rice centromeres contain only a minimal amount of satellite repeats, which allowed complete sequencing of these centromeres. Several active genes were found in these rice centromeres, representing the first true “centromeric genes” reported in any eukaryotes. The centromere of the maize B chromosome also presents a special model system for centromere research. The B centromere can be cytologically tracked in the maize genetic background, whereas individual centromeres are difficult to study cytologically in most eukaryotes. Numerous rearranged B centromeres have been developed, including inactivated and reactivated B centromeres, representing unique materials that are not available in other eukaryotes. The first generation of plant artificial chromosomes and engineered minichromosomes has also been developed.

The plant research community has generated a tremendous amount of information on the structure, function, and evolution of centromeres in several plant species during the last twenty years. Nevertheless, there has been no book and no special issue of any scientific journal that is dedicated to plant centromere research. This book includes a total of fourteen chapters that cover classical and modern centromere research in several plant species. It will be a valuable reference book or handbook for all plant scientists working on plant genome research. It can also be used as a reference book or textbook for upper level college classes with a theme on cytogenetics or genome analysis.

Jiming Jiang
Campbell-Bascom Professor

Department of Horticulture
University of Wisconsin–Madison

James A. Birchler
Curators' Professor of Biological Sciences
Division of Biological Sciences
University of Missouri

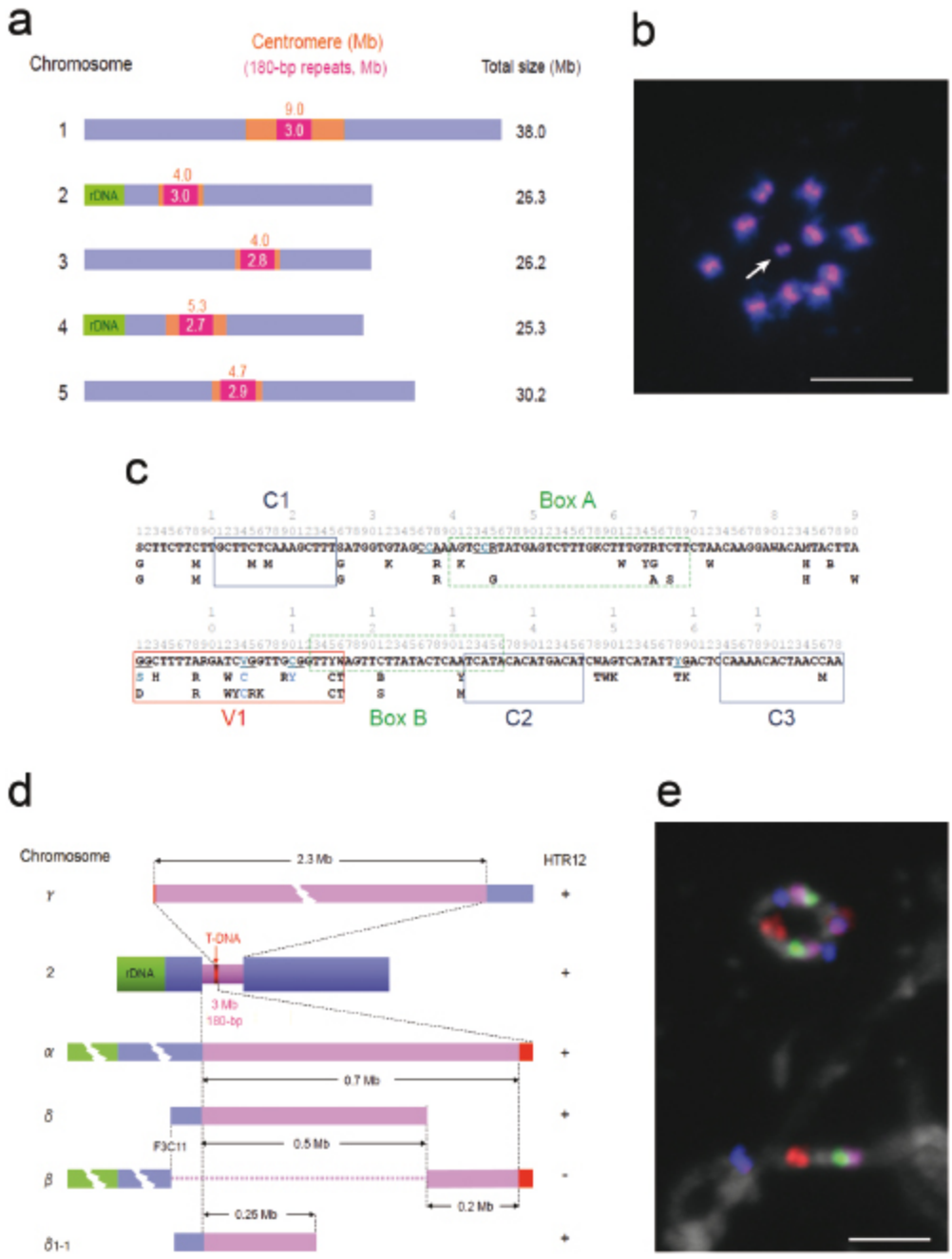
1

Arabidopsis Centromeres

Minoru Murata

Arabidopsis thaliana (L.) Heynh. is an annual flowering plant belonging to the family Brassicaceae. Since it has quite a small genome size and low amount of repetitive DNA sequences (see Meyerowitz, 1992, for the early history of the genome size estimation), it has become a model for molecular biological studies. Hence, its genome was the first among plant species to be sequenced (Arabidopsis Genome Initiative, 2000). This species has five pairs of chromosomes ($2n = 2x = 10$; Figure 1.1a), which is less than the chromosome number possessed by closely related species such as *A. lyrata* ($2n = 2x = 16$) and *A. arenosa* (*Cardaminopsis arenosa*; $2n = 2x = 16$). *A. suecica* ($2n = 4x = 26$) is an allotetraploid between *A. thaliana* and *A. arenosa* (Jakobsson et al., 2006).

Figure 1.1 (a) Schematic representation of the chromosome and centromere sizes of *Arabidopsis thaliana* (after Hosouchi et al., 2002). Orange box: genetically defined centromeric region; pink box: the central domain. (b) FISH image of a somatic prometaphase cell of *A. thaliana* ($2n = 10$ + miniα), probed with the 180-bp repeats. Arrow indicates a minichromosome (miniα; Murata et al., 2008) carrying truncated 180-bp repeat array. Bar = 5 μm. (c) Consensus sequences of 178-bp repeats from 41 ecotypes (upper), Columbia (Col; middle) and Col-edge (lower). Blue, red, and green dotted boxes indicate conserved (C1, C2, and C3), variable (V1) regions (Hall et al., 2003), and conserved Box A and B (Heslop-Harrison et al., 1999), respectively. Cytosine residues of underlined nucleotides in light blue are possibly differentially methylated depending on the centromeric or pericentromeric location (Zhang et al., 2008). (d) Schematic representation of chromosome 2 and its derivatives, showing centromere sizes and HTR12 (CENH3) localization, based on our previous data (Murata et al., 2008; Yokota et al., 2011). (e) FISH image of a pachytene cell of *A. thaliana*, probed with four different BAC clones mapped on the short arm of chromosome 2. Upper: miniΔ; lower: chromosome 2. Bar = 1 μm.



Chromosome size, which is highly related to genome size, has made cytological analysis difficult in *Arabidopsis* species. Nevertheless, it is very surprising that the first accurate report regarding the chromosome number ($2n = 10$ for *A. thaliana*) was made in 1907 (Laibach, 1907). Although the properties that made this plant suitable for genetic studies have been recognized for more than half a century (cf. Redei, 1992), the cytogenetical approach had

been quite limited until Sears's work (Steinitz-Sears, 1963; Sears and Lee-Chen, 1970). They assumed that the centromeres are located in or adjacent to the heterochromatic regions. Ambros and Schweizer (1976) applied Giemsa C-banding and confirmed that the centromeric regions of all chromosomes are heterochromatic. However, no DNA components of the centromeres had been revealed for a decade.

Centromere DNA structure

Regarding the centromeric DNA of *A. thaliana*, the first report was made by Martinez-Zapater and others (1986), which was followed by the work of Simoens and others (1988). Both research groups identified the same tandem repeat family, the unit size of which is approximately 180 bp (178~180 bp) and which constitutes approximately 0.8%–1.4% of the genome, among *Hin*dIII-digested DNA and the cosmid DNA library. The ladder pattern obtained via partial genome digestion by Southern blot analysis implied that the repetitive DNA sequences are arrayed in tandem. Although the former researchers speculated that the "180-bp family" lies within the heterochromatic blocks associated with centromeres or nucleolar organizing regions (Martinez-Zapater et al., 1986), neither research group could perform cytological analysis, due to the technical difficulty associated with the small size of chromosomes. Confirmation of the centromeric localization under microscopy had to wait for the establishment of the fluorescence in situ hybridization (FISH) technique. Using pAL1 as a probe, Maluszynska and Heslop-Harrison (1991) performed FISH and found that the FISH signals colocalize with the centromeric heterochromatin that could be visualized by DAPI-staining. A similar observation was made on mitotic metaphase cells using their own isolated two

repetitive DNA sequences (pAtMr1 and pAtHr1) having high homology to pAL1 (Murata et al., 1994; see Figure 1.1b as an example). In addition, they extended their observation to the meiotic chromosomes (prophase I to metaphase I) and noted that the FISH signals preferentially appeared at a limited part of heterochromatic regions, that is, within the heterochromatic blocks that are extended well at zygotene to pachytene stages.

The pAL1-family repetitive DNA sequences were reported to be tandemly arrayed to form large clusters of more than 50 kb (Martinez-Zapater et al., 1986). Pulsed-field gel electrophoresis revealed that the centromere clusters exceeded 1 Mb (Murata et al., 1994). Similarly, the use of different restriction enzymes that are insensitive to cytosine methylation allowed Round et al. (1997) to report that the 180-bp repeats form large clusters up to 1 Mb and that large (>400 kb) restriction fragments containing 180-bp repeat arrays total over 3 Mb in length in ecotype Columbia. They also indicated that there are size polymorphisms in the 180-bp repeat arrays between two ecotypes, Columbia and Landsberg *erecta*, which made it possible to map the 180-bp repeat arrays in the *Arabidopsis* genetic map (Round et al., 1997).

Copenhaver and others (1999) conducted a more extensive and accurate mapping of the centromeres and succeeded in connecting the centromeric contigs to the physical maps. In addition to the 180-bp repeat family, some other repeats such as 106A that have homology to the *Athila* retrotransposon were found to localize at the centromeric regions (Thompson et al., 1996; Brandes et al., 1997), but their participation in centromere function has not been demonstrated.

The genome project of *A. thaliana* was completed in December of 2000, and the 115.4-Mb region of the genome was recorded (Arabidopsis Genome Initiative, 2000). In the

genome project, over 5 Mb of centromeric regions and over 3 Mb of repetitive arrays (the 180-bp repeats and 5S rDNA) were sequenced, and the results showed that the centromeric regions are rich in various kinds of repetitive DNA sequences similar to those of many higher eukaryotes. However, the core regions within the centromeres, consisting mainly of the homogeneous 180-bp repeats, remain unrecorded. This high homogenization of the repeats with the head-to-tail repeat unit organization has made it difficult to find landmarks within the sequences. It was reported that 95% of the nucleotides are conserved, and that there is 99% conservation in the two boxes 30- and 24-bp long (Heslop-Harrison et al., 1999; Heslop-Harrison et al., 2003). However, these two boxes were not highly conserved across 41 ecotypes (Hall et al., 2003), and instead three other conserved regions (C1, C2, and C3) with 95% conservation and one variable region (V1) were noted (Figure 1.1c).

Based on the molecular and cytogenetical analyses of the centromere of chromosome 1, Haupt and others (2001) first estimated the centromere sizes of all five chromosomes, ranging from 1.4 Mb (Chromosome 3) to 2.3 Mb (Chromosome 1). Since there were still large gaps uncovered with existing BAC clones in the middle of the centromeres, the overall organization of the centromeres was investigated by restriction analysis of large DNA fragments (Kumekawa et al., 2000, 2001; Hosouchi, 2002). As a result, genetically defined centromeric regions were determined to range from 4.0 to 9.0 Mb, while the sizes of the central domains composing the 180-bp repeats were found to be close to one another in the range 2.7 to 3.0 Mb (Figure 1.1a).

Cytosine methylation and heterochromatin

Cytological studies have shown that the centromeric regions of *Arabidopsis* chromosomes are heterochromatic (Sears and Lee-Chen, 1970) and stain deeply with DAPI (Maluszynska and Heslop-Harrison, 1991). Since the DNA of constitutive heterochromatin is known to be highly methylated on cytosines, the centromeric repetitive DNA sequences have also been thought to be methylated. The highly methylated status of the 180-bp repeats has been indicated since the first discovery of the repeats (Martinez-Zapater et al., 1986). The discovery was based on the use of the restriction enzymes *Hpa*II and *Msp*I, both of which recognize 5′-CCGG-3′, and the former is sensitive and latter insensitive to the second cytosine methylation. Although asymmetrical cytosine methylations are also common in the centromeric repeats and not all repeat units contain the 5′-CCGG-3′ sequence, this kind of symmetrical cytosine methylation has been used to screen the hypomethylation mutants in *A. thaliana* (Vongs et al., 1993).

Various approaches have been used to elucidate the relationship between the centromere, heterochromatin, and cytosine methylation as well as histone methylation (e.g, Luo et al., 2004). One of the most important findings regarding *Arabidopsis* centromere structure and functions concerns hypomethylation on the core regions of the centromeres, which are parts of the 180-bp repeat arrays and predominantly covered with the centromere-specific histone H3 (CENH3, HTR12, or CENP-A homologous in *A. thaliana*; Zhang et al., 2008). Using anti-5-methylcytosine antibody, it was shown that the 180-bp repeats associated with CENH3, which were referred to as the CEN chromatins, are distinctly hypomethylated, whereas the same repeat family in the pericentromeric heterochromatin is heavily

methylated, and histone H3 dimetylated at lysine 9 (H3K9me2) is significantly reduced in the DNA-hypomethylated centromere regions. This differentiation in methylation status between the centromeric and pericentromeric regions might be related to differences in DNA sequence of the 180-bp repeats analyzed (Hall et al., 2003; Figure 1.1c). Since the CEN chromatins are flanked by heterochromatin enriched with H3K9me2, this situation is very similar to that in *S. pombe* (Partridge et al., 2000) and in *D. melanogaster* (Blower et al., 2002), although no DNA methylation is involved in *S. pombe*. DNA methylation and/or DNA-methylation-associated H3K9me2 or other histone modifications were suggested to act as a boundary to isolate the CEN chromatin (Zhang et al., 2008). In addition to the boundary role, heterochromatin at the pericentromeric regions could have additional roles in recruiting cohesin for sister chromatid cohesion (Gartenberg, 2009).

Centromere proteins

The centromere is a multifunctional complex, involving kinetochore formation, sister chromatid adhesion and separation, microtubule attachment, chromosome movement, heterochromatin establishment, and mitotic checkpoint control. Among these functions, kinetochore formation is the most fundamental and essential. There are more than 60 constituent proteins of kinetochores in budding yeast (McAinsh et al., 2003), and more than twenty of these kinetochore proteins are conserved from yeasts to mammals (Amor et al., 2004; Table 1.1). This conservation is in striking contrast to the poor conservation of centromere DNA sequences (Henikoff et al., 2001).

Table 1.1 Centromere proteins of *A. thaliana* and four other species

Localization (Network)*	Species					
	H. sapiens	*D. melanogaster*	*S. pombe*	*S. cerevisiae*	*A. thaliana*	**
IC	INCENP	INCENP	Pic1	Sli15	AtINCENP	1
IC	AuroraB	IPLI	Ark1	Ipl1	AtAUR3	2
IC	CENP-A	CID	Cnp1	Cse4	HTR12	3
IC	CENP-B	—	Abp1, Cbh1,2	—	—	
IK (CCAN)	CENP-C	CENP-C	Cnp3	Mif2	AtCENP-C	4
IK (CCAN)	CENP-H		Fta3	Mcm16		
IK (CCAN)	CENP-I		Mis6	Ctf3		
IK (CCAN)	CENP-K		Sim4			
IK (CCAN)	CENP-L		Fta1			
IK (CCAN)	CENP-M		Mis17	Iml3		
IK (CCAN)	CENP-N		Mis15			
IK (CCAN)	CENP-O		Mal2	Mcm21	AtCENP-O	5
IK (CCAN)	CENP-P		Fta2			
IK (CCAN)	CENP-Q		Fta7			
IK (CCAN)	CENP-R					
IK (CCAN)	CENP-U					
IK (CCAN)	CENP-S			YOL86-A		
IK (CCAN)	CENP-X					
IK (CCAN)	CENP-T		SpBC800			
IK (CCAN)	CENP-W					
OK (KMNN)	hMis12	CG18156	Mis12	Mtw1	AtMIS12	6
OK (KMNN)	DSN1		Dsn1/Mis13	Dsn1		
OK (KMNN)	NNF1	CGI13434	Nnf1	Nnf1		
OK (KMNN)	NSL1	CGI1558	Nsl1/Mis14	Nsl1		
OK (KMNN)	Hec1/NDC80	Ndc80	Ndc80	Tid3	AtNDC80	7
OK (KMNN)	NUF2	Nuf2	Nuf2	Nuf2	AtNUF2	8
OK (KMNN)	SPC24	(GI12063)	Spc24	Spc24		
OK (KMNN)	SPC25	CG7242	Spc25	Spc25	AtSPC25	9
OK (KMNN)	KNL1	CG11451	Spc7	Spc105		
(facultative)	CENP-E	CENP-meta	Tea2	Kip2	AtCENP-E	10
(facultative)	CENP-F	Spn		Atg11		
(facultative)	CENP-V					
(CACE)	HJURP	(CAL1?)	YK12	Scm3		
(CACE)	Mis18α		Mis18			
(CACE)	Mis18β		Mis18			
(CACE)	M18BP1		Mis18			
(CACE)	RbAp48	RbAp48	Mis16	Msi1	AtMSI1	11
(CACE)	RbAp46		Mis16	Msi1		

*IC = inner centromere; IK = inner kinetochore; OK = outer kinetochore; CACE = CENP-A chromatin establishment; CCAN = constitutively centromere-associated network; KMNN = KNL1/Mis12 complex/Ndc80 complex network.

**Loci and references: 1. AT5G55820; Kirioukhova et al., 2011. 2. AT2G45490; Kurihara et al., 2006. 3. AT1G01370; Talbert et al., 2002. 4. AT1G15560; Ogura et al., 2004; Talbert et al., 2004. 5. AT5G10710; Direct submission to TAIR database, Swarbreck et al., 2011. 6. AT5G35520; Sato et al., 2005. 7. AT3G54630. 8. AT1G61000. 9. AT3G48210. 10. AT2G21380. 11. AT5G58230; Direct submission to TAIR database, Swarbreck et al., 2011.

Although studies on kinetochore proteins have been performed mainly in yeasts and mammals, some of the plant counterparts have been identified since the pioneering work on maize CENP-C (Dawe et al., 1999). In *A. thaliana*, Talbert and colleagues (2002) first identified the HTR12 protein as a centromere-specific histone H3 variant

(CENH3), which corresponds to CENP-A in mammals. This report certainly accelerated subsequent centromere studies, since CENP-A or its orthologues are present in all eukaryotes that have been investigated to date, and are only detected on functional centromeres (Warburton et al., 1997). Interestingly, HTR12 is detected on all centromeres in *A. suecica* (allotetraploid, $2n = 4x = 26$) and *A. thaliana* ($2n = 2x = 10$) but not in *A. arenosa* ($2n = 2x = 16$) that is another parent of *A. suecica*. This suggests a unique evolutionary force important for the centromere proteins. The close interaction of HTR12 with the 180-bp repeats was shown by the chromatin immunoprecipitation (ChIP) assay, but an interaction with *Athila*, a *Ty3*/gypsy-type retroelement, was not detected (Nagaki et al., 2003).

A gene (AT2G06660) encoding CENP-B-like protein was thought to exist in the *Arabidopsis* genome, but this is now doubtful since its homology to CENP-B of mammals and Abp1, Cbh1, and Cbh2 of fission yeast is unclear, and no distinct transcription and/or translation from the CENP-B-like gene has been confirmed (Murata, 2002). The *Arabidopsis* counterpart of CENP-C (AtCENP-C) was identified based on the homology to DNA sequences of maize CENP-C (Ogura et al., 2004; Talbert et al., 2004). Human CENP-C is one of the few centromere proteins having DNA-binding ability, and its close association to CENP-A has been suggested (Perpelescu and Fukagawa, 2011). Although the C-terminal amino acid sequence of AtCENP-C was conserved among plant species, no similarity to animal or fungal CENP-Cs was found, except for the CENP-C motif (Talbert et al., 2004).

Mis12 was first identified as one of the kinetochore proteins in *S. pombe* (Goshima et al., 1999), and its human orthologue was shown to be a component of the Mis12/MIND complex comprising Mis12, Dsn1, Nnf1, and Nsl1 (Perpelescu and Fukagawa, 2011). Despite the poor overall similarity to fission yeast and human Mis12, Goshima

and colleagues (2003) predicted the Mis12 homologue in *A. thaliana* using Block Maker (Henikoff et al., 1998) and MAST (Bailey and Gribskov, 1998) analysis. The centromere localization of the putative AtMIS12 was confirmed by immunostaining with the antibody raised against a peptide synthesized from the putative amino acid sequence (Sato et al., 2005).

For other kinetochore proteins, orthologues have not been identified in *Arabidopsis* until recently, mainly due to the rapid findings of novel kinetochore proteins in humans and yeasts and their poor homologies to plant orthologues. Very recently, however, six counterparts were identified based on InterPro domain analysis (D. Li, personal communication) and added to the TAIR database (http://arabidopsis.org). To date, 11 centromere proteins have been listed in *A. thaliana* (Table 1.1), although the centromere localization and function of the newly-added proteins have not yet been revealed. In the inner centromere structure, three of four components except CENP-B have been identified among human, fly, fission and budding yeasts, and *Arabidopsis*. Since CENP-B or its homologues have been shown to be inessential in mice and fission yeasts (Kapoor et al., 1998; Perez-Castro et al., 1998; Baum and Clarke, 2000), it is not surprising that no CENP-B counterparts have been detected in *Arabidopsis* or other eukaryotes. This fact suggests that the inner centromere structure is conserved well from yeasts to animals and plants. Similarly, the structure of the outer kinetochore seems conserved among the eukaryotes, since most of the constituent protein counterparts have been identified, even in *Arabidopsis* (four of nine counterparts). On the other hand, it is difficult to determine the components of the inner kinetochore in *Arabidopsis*, except AtCENP-C and -O. Although a group of those components, called the constitutive centromere-associated network, are conserved in vertebrates, these orthologues

have seldom been identified in *D. melanogaster* or *C. elegans* (Perpelescu and Fukagawa, 2011). For example, the CENP-H/I complex was shown to be necessary for centromere-targeting of newly-synthesized CENP-A (Okada et al., 2006), but in *A. thaliana,* the CENP-I/Mis6 homologues remain unidentified (Sato et al., 2005). These data suggest the possibility that plants, as well as some invertebrates, have different kinetochore structures from those of vertebrates, and this idea is supported by the finding that the classical tri-layer structure of vertebrate kinetochores has not been detected in plants (Wilson, 1968; Dawe et al., 2005).

CENP-A or CENH3 is a key protein that interacts with centromeric DNA sequences (Henikoff et al., 2001). Its necessity for kinetochore assembly was first shown in mouse null mutants for Cenpa (Howman et al., 2000), and was also confirmed in *A. thaliana* using its tetraploid plants (Ravi et al., 2010; please see Chapter 13 for details). Therefore, it is very important to know the process of CENP-A chromatin establishment for kinetochore formation, which is divided into centromere priming, CENP-A uploading, and maintenance (Perpelescu and Fukagawa, 2011). In the process, three to five components have been identified in humans and fission yeasts (Table 1.1). Among them, HJURP is the most important component, working as a CENP-A-specific chaperone. In *Drosophila*, however, it has just been reported that CAL1 (Chromosome ALignment defect 1), whose amino acid sequence has diverged from that of HJURP and its yeast counterparts, has similar functions to HJURP and Scm3 (Mellone et al., 2011). This sort of divergence might make it difficult to determine the HJURP/Scm3 counterpart in *Arabidopsis*.

Functional domains

As described above, the *Arabidopsis* centromeric regions are preferentially occupied by the “180-bp repeat” family. Since the array size of the 180-bp cores has been estimated to be about 2.7–3 Mb for all five chromosomes (Kumekawa et al., 2000, 2001; Hosouchi et al., 2002), this size seems important for centromere functionality and accurate chromatid segregation during cell division. However, ChIP assays suggested that only subsets of the 180-bp repeat arrays are involved in centromere function (Nagaki et al., 2003). More direct evidence was obtained from chromatin-fiber immunolabeling and the FISH technique, which demonstrated that HTR12 proteins localize only on a limited number of copies of the 180-bp repeats (Shibata and Murata, 2004).

Minichromosomes with truncated centromeres are quite useful for elucidating the relationship between the size of repeat arrays and functionality, as shown in fruit fly (Sun et al., 2003) and humans (Spence et al., 2002). In *A. thaliana*, several minichromosomes have been isolated (Table 1.2). Since most of these are relatively stable and transmissible to the next generation, they are maintained as partial trisomic lines. All of these minichromosomes were found to carry a shorter array of the centromeric satellite, and they are valuable for analyzing centromere function (Murata et al., 2006; Murata et al., 2008; Yokota et al., 2011). The minichromosome mini4S was found in progeny of telotrisomic Tr1A plants of Landsberg *erecta* and was shown to have originated from the short arm of chromosome 4 and possesses a truncated centromere (Murata et al., 2006). This “mini4S,” the size of which was estimated to be approximately 7.5 Mb, contains only about 1 Mb, or about one-third of the amount of centromeric 180-bp repeats in the normal chromosome 4. However, it is relatively stable at mitosis, particularly in the Columbia background, and the transmission rate to the next generation was comparable to