Progress in Molecular and Subcellular Biology

Đurđica Ugarković
Editor

Centromere

Structure and Evolution

Springer

Progress in Molecular and Subcellular Biology

48

Volumes Published in the Series

Progress in Molecular and Subcellular Biology

Subseries: Marine Molecular Biotechnology

Volume 33
Silicon Biomineralization
W.E.G. Müller (Ed.)

Volume 34
Invertebrate Cytokines and the Phylogeny of Immunity
A. Beschin and W.E.G. Müller (Eds.)

Volume 35
RNA Trafficking and Nuclear Structure Dynamics
Ph. Jeanteur (Ed.)

Volume 36
Viruses and Apoptosis
C. Alonso (Ed.)

Volume 37
Sponges (Porifera)
W.E.G. Müller (Ed.)

Volume 38
Epigenetics and Chromatin
Ph. Jeanteur (Ed.)

Volume 39
Echinodermata
V. Matranga (Ed.)

Volume 40
Developmental Biology of Neoplastic Growth
A. Macieira-Coelho (Ed.)

Volume 41
Molecular Basis of Symbiosis
J. Overmann (Ed.)

Volume 42
Antifouling Compounds
N. Fusetani and A.S. Clare (Eds.)

Volume 43
Molluscs
G. Cimino and M. Gavagnin (Eds.)

Volume 44
Alternative Splicing and Disease
Ph. Jeanlevr (Ed.)

Volume 45
Asymmetric Cell Division
A. Macieira Coelho (Ed.)

Volume 46
Marine Toxins as Research Tools
N. Fusetani and W. Kem (Eds.)

Volume 47
Biosilica in Evolution, Morphogenesis, and Nanobiotechnology
W.E.G. Müller and M.A. Grachev (Eds.)

Volume 48
Centromere
Đurđica Ugarković (Ed)

Đurđica Ugarković
Editor

Centromere

Structure and Evolution

Springer

Editor
Đurđica Ugarković
Ruder Boskovic Institute
Center for Marine Research
Bijenicka 54
HR-10001 Zagreb
P.O. Box 1016
Croatia

ISSN 0079-6484
ISBN 978-3-642-00181-9 e-ISBN 978-3-642-00182-6
DOI 10.1007/978-3-642-00182-6

Library of Congress Control Number: 2008944103

Cover design: Boekhorst Design BV, The Netherlands

Printed on acid-free paper

springer.com

Preface

The centromere is a chromosomal region that enables the accurate segregation of chromosomes during mitosis and meiosis. It holds sister chromatids together, and through its centromere DNA–protein complex known as the kinetochore binds spindle microtubules to bring about accurate chromosome movements. Despite this conserved function, centromeres exhibit dramatic difference in structure, size, and complexity. Extensive studies on centromeric DNA revealed its rapid evolution resulting often in significant difference even among closely related species.

Such a plasticity of centromeric DNA could be explained by epigenetic control of centromere function, which does not depend absolutely on primary DNA sequence. According to epigenetic centromere concept, which is thoroughly discussed by Tanya Panchenko and Ben Black in Chap. 1 of this book, centromere activation or inactivation might be caused by modifications of chromatin. Such acquired chromatin epigenetic modifications are then inherited from one cell division to the next. Concerning centromere-specific chromatin modification, it is now evident that all centromeres contain a centromere specific histone H3 variant, CenH3, which replaces histone H3 in centromeric nucleosomes and provides a structural basis that epigenetically defines centromere and differentiates it from the surrounding chromatin. Recent insights into the CenH3 presented in this chapter add important mechanistic understanding of how centromere identity is initially established and subsequently maintained in every cell cycle.

To explain contradiction between rapid evolution of centromeric DNA and centromeric histones on one site and conservation of centromere function on the other one, a model termed "centromere drive" has been proposed by Steven Henikoff and Harmit Malik in 2002. According to this model, asymmetry in female meiosis acts as a driving force in centromere evolution by inducing a constant genetic conflict between two essential genetic elements: centromeric satellite DNA and centromeric histones or other satellite-binding proteins. Such a conflict is responsible for rapid centromere evolution. In Chap. 2 of this book, Harmit Malik summarizes the evidence in favor of the centromere-drive model and its implications for centromere evolution.

Although extant data favor centromere being epigenetic structure, it is also clear that centromere formation is based on DNA, in particular tandemly repeated satellite DNA, which is a predominant component of many centromeres. Presence of conserved structural motifs within satellite DNAs indicates existence of structural

determinants which are prerequisite for centromere function. In Chap. 3, Đurđica Ugarković discusses the role of DNA in centromere establishment and proposes that centromere is formed from adapted sequences with certain structural characteristics. After exaptation, that is, after becoming functional, these sequences can reside within the genome for long evolutionary periods and create so called *satellite* DNA *library*.

Recently, it is revealed that centromeres are transcriptionally active and RNA is identified as a structural component of kinetochore, essential for centromere function. In Chap. 4, Rachel O'Neill and Dawn Carone highlight the current understanding of centromere structure and evolution, as well as role of transcription in centromere function, using as a model system marsupials. Because of small size and importance in speciation, marsupial centromere represents a valuable mammalian centromere model.

Neocentromere formation and evolution of new centromeres have been thoroughly discussed by Rocchi, Stanyon, and Archidiacono in Chap. 5. Using primates as a model system, they explain mechanisms leading to the formation of both types of centromeres and define centromere forming domains that preserve features that trigger neocentromere emergence over tens of millions of years of evolutionary time. Findings described in this chapter reveal that centromeres can origin, live, and go extinct, but inactive and ancient ones can be also "reused" as centromere seeding points in evolution.

Intense investigation of centromere components, DNA, and proteins has been performed in different plant species, in particular Arabidopsis and Gramineae, during this decade. A comprehensive review written by Jiang and Murata with collaborators summarizes (Chap. 6 in this book) present data on plant centromere components. In addition, evolution of plant centromere is discussed as well as future directions in plant centromere investigation.

Contents

Contributors

Nicoletta Archidiacono
Dipartimento di Genetica e Microbiologia, Via Amendola 165/A,
70126 Bari, Italy

Ben E. Black
Department of Biochemistry and Biophysics, University of Pennsylvania,
Philadelphia, PA 19104-6059, USA
blackbe@mail.med.upenn.edu

Dawn M. Carone
Center for Applied Genetics and Technology, Department of Molecular
and Cell Biology, University of Connecticut, Storrs, CT 06269, USA

Cory Hirsch
Department of Horticulture, University of Wisconsin-Madison, Madison,
WI 53706, USA

Jiming Jiang
Department of Horticulture, University of Wisconsin-Madison, Madison,
WI 53706, USA
jjiang1@wisc.edu

Harmit S. Malik
Division of Basic Sciences, Fred Hutchinson Cancer Research Center,
1100 Fairview Avenue N, Seattle, WA 98109, USA
hsmalik@fhcrc.org

Minoru Murata
Research Institute for Bioresources, Okayama University
Kurashiki 710-0046, Japan
mmura@rib.okayama-u.ac.jp

Kiyotaka Nagaki
Research Institute for Bioresources, Okayama University
Kurashiki 710-0046, Japan

Rachel J. O'Neill
Center for Applied Genetics and Technology, Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA
rachel.oneill@uconn.edu

Tanya Panchenko
Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104-6059, USA

Mariano Rocchi
Dipartimento di Genetica e Microbiologia, Via Amendola 165/A, 70126 Bari, Italy
rocchi@biologia.uniba.it

Roscoe Stanyon
Dipartimento di Biologia Evoluzionistica, Via del Proconsolo 12, 50122 Firenze, Italy

Đurđica Ugarković
Department of Molecular Biology, Ruđer Bošković Institute, Bijenička 54, HR-10002 Zagreb, Croatia
ugarkov@irb.hr

Jason Walling
Department of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA

Chapter 1
The Epigenetic Basis for Centromere Identity

Tanya Panchenko and Ben E. Black

Contents

T. Panchenko and B.E. Black (✉)
Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104-6059
e-mail: blackbe@mail.med.upenn.edu

Đ. Ugarković (ed.), *Centromere,* Progress in Molecular and Subcellular Biology 48,
DOI: 10.1007/978-3-642-00182-6_1, © Springer-Verlag Berlin Heidelbarg 2009

Abstract The centromere serves as the control locus for chromosome segregation at mitosis and meiosis. In most eukaryotes, including mammals, the location of the centromere is epigenetically defined. The contribution of both genetic and epigenetic determinants to centromere function is the subject of current investigation in diverse eukaryotes. Here we highlight key findings from several organisms that have shaped the current view of centromeres, with special attention to experiments that have elucidated the epigenetic nature of their specification. Recent insights into the histone H3 variant, CENP-A, which assembles into centromeric nucleosomes that serve as the epigenetic mark to perpetuate centromere identity, have added important mechanistic understanding of how centromere identity is initially established and subsequently maintained in every cell cycle.

1.1 Introduction

Mitotic segregation of the genome is an essential process for all eukaryotes, and all eukaryotic chromosomes use a control locus – the centromere – to self-direct their own segregation. It has been clear for decades that the underlying DNA sequences at centromeres are highly divergent, while the genes found along the chromosome arms are highly conserved. Ten years ago, the characterization of human neocentromeres laid bare a true paradox at the centromere: while megabase arrays of repetitive DNA are typically found at eukaryotic centromeres, the repeats themselves are neither required for centromere identity nor for centromere function (Eichler 1999). Function, in the case of centromeres, is defined as the ability of the locus to build a kinetochore at meiosis and mitosis that serves as the physical connection of the chromosome to the microtubule-based spindle. Given the central nature that centromeres play in directing inheritance in the germline and in preserving genome integrity in somatic cells, the resolution of this paradox has emerged as a key problem in biology. Many lines of evidence point to strong epigenetic mechanisms to determine centromere identity. As in the case of epigenetic mechanisms that modulate gene expression, the studies of centromere epigenetics have focused on chromatin structure.

The specific architecture and scale of an individual centromeric chromatin domain can vary substantially between divergent eukaryotic species. All functional centromeres, however, contain a centromere-specific histone H3 variant (CENP-A from humans was the first centromere-specific histone to be identified (Earnshaw et al. 1986; Earnshaw and Cooke 1989; Earnshaw and Rothfield 1985; Palmer et al. 1987, 1989, 1991; Sullivan et al. 1994)). Centromere identity is typically defined by the presence of an array of nucleosomes in which CENP-A replaces H3. Data from several diverse model systems, including some yeast systems (where there is a stronger genetic component than in metazoans), have greatly contributed to our current understanding of how centromeres are specified. In this review, we survey some of the classic studies from diverse eukaryotic species that have each shaped our current view of the centromere as an epigenetic locus and discuss recent studies that have advanced our understanding of centromere identity and function.

1.2 The Budding Yeast Centromere

The *Saccharomyces cerevisiae* centromere is the most thoroughly characterized centromere in any system. It is an extreme example of a centromere due to its very small size (125 bp) and strong DNA sequence dependence. The simplicity of the *S. cerevisiae* centromere and the tractable genetics of the organism have led to elegant experiments that have elucidated its nature. While the budding yeast centromere could be viewed as an exception to the rule of epigenetic centromere formation, a discussion of its well understood key features will put findings from other eukaryotic species into context.

1.2.1 Genetic Definition of a Centromere

The identification of a region on chromosome III that is required for centromere function provided the first indication that the centromeres of budding yeast are defined genetically (Clarke and Carbon 1980). This isolated sequence imparts mitotic and meiotic stability to circular plasmids (which also carry an autonomously replicating sequence (ARS) to confer replication in S-phase), functionally creating a mini-chromosome (Stinchcomb et al. 1979; Clarke and Carbon 1980; Fitzgerald-Hayes et al. 1982a;). The minichromosome/chromosome stability approach was extended to identify the functional centromeres on each of the budding yeast chromosomes, and eventually generated a centromere consensus sequence of 125 bp (Clarke and Carbon 1980; Fitzgerald-Hayes et al. 1982a, b; Panzeri and Philippsen 1982; Stinchcomb et al. 1982; Hieter et al. 1985; Maine et al. 1984; Neitz and Carbon 1985; Mann and Davis 1986; Cottarel et al. 1989). This consensus sequence is clearly comprised of three parts, centromere DNA element I, II, and III (CDE I, CDE II, and CDE III), each of which serves distinct roles and are made up of unique sequences (Fitzgerald-Hayes et al. 1982b; Hieter et al. 1985; Neitz and Carbon 1985). CDE I and CDE III represent the right and left boundaries of the centromeric DNA region and are most conserved from chromosome to chromosome. The AT-rich CDE II (>90% AT) is relatively invariant in nucleotide composition and length but varies widely in sequence (Fitzgerald-Hayes et al. 1982b; Hieter et al. 1985). Mutational analysis of these regions indicates that CDE II and CDE III are most sensitive to variations (Carbon and Clarke 1984; Fitzgerald-Hayes 1987; Gaudet and Fitzgerald-Hayes 1987; Murphy and Fitzgerald-Hayes 1990; Murphy et al. 1991). Decreasing the AT content as well as altering the length of CDE II decreases plasmid stability in mitosis by ~1,000 fold, whereas individual point mutations are tolerated. On the other hand, the CDE III element is most sensitive to individual point mutations (McGrew et al. 1986; Cumberledge and Carbon 1987; Jehn et al. 1991). The highly sequence-conserved CDE I and CDE III elements recruit the Cbf1p homodimer and CBF3 protein complex, respectively, via sequence-specific DNA binding protein modules (Bram and Kornberg 1987; Baker et al. 1989; Cai and Davis 1989; Jiang and Philippsen 1989;

Lechner and Carbon 1991). CBF1 is dispensable for kinetochore function, whereas the CBF3 complex plays an essential role in building a functional budding yeast kinetochore (Goh and Kilmartin 1993; Sorger et al. 1994).

1.2.2 A Single Specialized Nucleosome at the Centromere

Also present at the 125 bp centromere of budding yeast is a single nucleosome containing the CENP-A relative, *Sc*CENP-A (also called Cse4p) (Stoler et al. 1995; Furuyama and Biggins 2007). The 125 bp sequence is ~20 bp shorter than what is required for full wrapping of a canonical histone octamer. It is not clear if some surrounding sequences are used to wrap a putative Cse4p-containing nucleosome or if it unwraps ~10 bp at each DNA entry/exit site. The fact that the centromeric chromatin that is protected by nuclease digestion extends 160–200 bp (Bloom and Carbon 1982; Funk et al. 1989) suggests that full wrapping of the *Sc*CENP-A-containing nucleosome is indeed possible. *Sc*CENP-A specifically associates with the centromeric regions of yeast chromosomes but not with other AT-rich regions in the genome (Meluh et al. 1998; Furuyama and Biggins 2007). Both centromere DNA and *Sc*CENP-A are required for centromere function and the gene encoding *Sc*CENP-A is essential for viability (Clarke and Carbon 1983; Stoler et al. 1995; Meluh et al. 1998).

The composition of the budding yeast centromeric nucleosome has been the topic of recent studies. An affinity purification of yeast centromeres suggests that the composition is similar to a canonical nucleosome (Fig. 1.1a), but with *Sc*CENP-A replacing both copies of histone H3 (Fig. 1.1b; i.e., two copies each of H2A, H2B, H4, and *Sc*CENP-A) (Westermann et al. 2003). Recent evidence to the contrary indicates that the *Sc*CENP-A-containing nucleosomes lack histones H2A and H2B

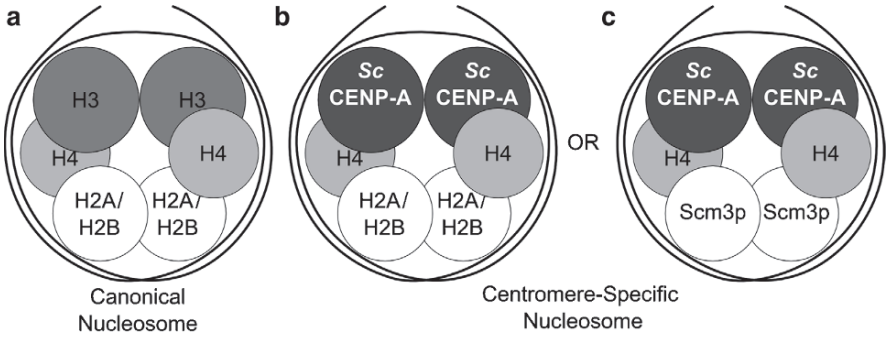


Fig. 1.1 Nucleosome composition in *Saccharomyces cerevisiae*. (**a**) Composition of canonical nucleosomes. (**b**) *Sc*CENP-A replaces H3 in centromere-specific nucleosomes. (**c**) Proposed centromeric nucleosome composition where Scm3p replaces H2A/H2B (Mizuguchi et al. 2007)

(Mizuguchi et al. 2007). Instead the centromeric protein, Scm3p (Camahort et al. 2007; Mizuguchi et al. 2007; Stoler et al. 2007), is bound to the (*Sc*CENP-A-H4)$_2$ tetramer at the centromere (Fig. 1.1c; (Mizuguchi et al. 2007)). Scm3p binds to the *Sc*CENP-A-H4 tetramer as a dimer, forming a hexamer. Scm3p has also been shown to be required for *Sc*CENP-A localization to the centromere as well as progression through cell cycle (Camahort et al. 2007; Stoler et al. 2007). Importantly, Scm3p interacts with Ncd10p (Camahort et al. 2007), a component of CBF3, providing a potential link between the CBF3 complex and the *Sc*CENP-A-containing nucleosome. Although CBF3 is required for localizing *Sc*CENP-A to the centromere (Measday et al. 2002; Ortiz et al. 1999), no direct physical interaction has been reported.

1.2.3 Alternative Segregation Mechanisms for the 2 μm Plasmid

In addition to chromosomal centromere sequences, at least one other type of DNA sequence may be segregated during budding yeast mitosis in a manner requiring *Sc*CENP-A. The 2-μm plasmids (parasitic entities that inhabit yeast cells) encode several proteins including Rep1 and Rep2 that are recruited to the plasmid's STB locus in a process that generates plasmid stability through mitosis (Jayaram et al. 1983, 1985; Kikuchi 1983; Som et al. 1988; Scott-Drew and Murray 1998). Even though these plasmids lack a centromere, and are not thought to form a functional kinetochore, the presence of *Sc*CENP-A-containing chromatin has been proposed to carry out some centromere function, such as mitotic regulation of duplicated plasmid cohesion (Hajra et al. 2006). It should also be noted that the STB locus of the 2-μm plasmid was found to confer mitotic stability in an early plasmid stability screen (Hieter et al. 1985).

Nevertheless, the formation of bona fide budding yeast centromeres is fundamentally a genetic process where the underlying DNA sequence dictates its identity. In this way the *S. cerevisiae* centromere is exceptional among other well studied eukaryotic species where centromere identity is specified epigenetically, as discussed later.

1.3 The Fission Yeast Centromere

Fission yeast, as compared to budding yeast, has evolved in a vastly different centromere structure and organization. The underlying DNA length is substantially increased, with the centromeric regions on each chromosome ranging from 30 to 100 kb (Clarke et al. 1986; Nakaseko et al. 1986; Fishel et al. 1988; Chikashige et al. 1989; Hahnenberger et al. 1989; Murakami et al. 1991). Fission yeast centromeres

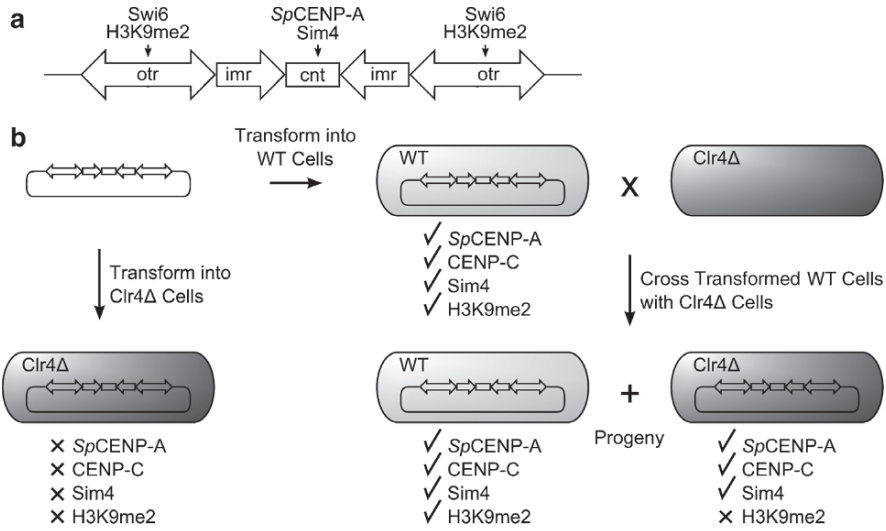


Fig. 1.2 Centromere formation on minichromosomes in *Schizosaccharomyces pombe*. (**a**) Diagram of the fission yeast centromere. (**b**) Experiment that demonstrates that initial formation of *Sp*CENP-A-containing nucleosomes on naked DNA templates requires pericentromeric heterochromatin. If centromere identity is initially established, however, the kinetochore-forming chromatin containing *Sp*CENP-A perpetuates in the absence of pericentromeric heterochromatin (Folco et al. 2008)

are arranged into clearly defined regions (Fig. 1.2a). The central region, *cnt*, is essential, non-repetitive, and flanked by two identical inverted repeats (*imr*) (Chikashige et al. 1989; Hahnenberger et al. 1989; Murakami et al. 1991; Takahashi et al. 1992; Clarke et al. 1993; Steiner et al. 1993; Steiner and Clarke 1994). *cnt* is found associated with the *Schizosaccharomyces pombe* orthologue of CENP-A (*Sp*CENP-A; also called Cnp1) and is the site of kinetochore formation (Takahashi et al. 2000). The central domain (*cnt* + *imr*) is in turn flanked by repetitive outer repeats (*otr*) of varying lengths (Clarke and Baum 1990; Hahnenberger et al. 1991; Kuhn et al. 1991; Polizzi and Clarke 1991; Steiner et al. 1993). Non-coding RNAs are transcribed from within *otr* and are processed into short duplex RNAs by Dicer (Volpe et al. 2002). These duplex RNAs, similar to siRNAs involved in regulating gene expression, bind to the argonaute protein (Ago1 in fission yeast) and ferry the Ago1-containing RITS complex to the centromere, all of which culminates in the formation of a pericentromeric heterochromatin domain that is required, along with kinetochore function from the *cnt* domain, for accurate chromosome segregation (Hall et al. 2002; Volpe et al. 2002; Motamedi et al. 2004; Verdel et al. 2004). The pericentromeric heterochromatin compartment is enriched in Swi6 (the fission yeast analog of Heterochromatin Protein 1 – HP1), depleted of acetylated histone H3 marks, and enriched for methylation on Lys9 of H3 (Ekwall et al. 1997; Nakayama et al. 2000, 2001; Partridge et al. 2000; Noma et al. 2001). All these marks indicate a silenced chromatin state and transcriptional reporter cassettes inserted within the pericentromeric heterochromatin are indeed silenced (Allshire et al. 1994, 1995; Partridge et al. 2000).

1.3.1 CENP-A-Containing Nucleosomes Epigenetically Mark Centromere Location

An early indication that epigenetic mechanisms may be employed to specify fission yeast centromeres came from the finding that there is no centromere sequence resembling the centromere sequences of budding yeast. While *Sp*CENP-A is essential for viability and centromere function, no particular DNA sequence is required for its assembly as the *cnt* domain is replaceable with a non-centromeric sequence without compromising centromere identity or continued *Sp*CENP-A loading (Castillo et al. 2007).

S. pombe has been proven to be a very informative model system for centromere biology due to its higher order domain structure and the powerful genetic tools that have been developed for fission yeast. Using traditional genetic screening, several components have been identified that are required for chromosome segregation, in general, and for *Sp*CENP-A centromere localization, in particular. Yanagida and colleagues have extensively screened for mutants showing minichromosome instability (Mis mutants). One of the Mis gene products, Mis6 (Saitoh et al. 1997), is found at the *cnt* region and is required for *Sp*CENP-A centromere localization (Takahashi et al. 2000). Sim4, which physically interacts with Mis6 in a complex that also contains Mis15 and Mis17 (Hayashi et al. 2004), was independently identified in a screen for mutants defective in centromeric transcriptional silencing and is also required for *Sp*CENP-A localization at centromeres (Pidoux et al. 2003). Two additional Mis proteins, Mis16 and Mis18, physically interact with each other and are both required for the centromeric localization of *Sp*CENP-A (Hayashi et al. 2004). Mis16 is the fission yeast orthologue of the human chromatin assembly factor 1 (CAF1) subunit p46/p48 (also known as RbAp46/48) that binds to H3/H4 tetramers and/or dimers via H4 contacts (Murzina et al. 2008). Mis16 may also associate with the corresponding sub-nucleosomal histone complex containing *Sp*CENP-A and H4 that is thought to exist prior to nucleosome assembly. Another potential component of the centromere chromatin assembly pathway is Ams2, a transcription factor from the GATA protein family, which is required for SpCENP-A localization (Chen et al. 2003). The expression of Ams2 is regulated with its levels peaking during the G1/S phases of the cell cycle, prior to S-phase when a burst of *Sp*CENP-A loading occurs (Takayama et al. 2008). A more recent genetic screening for mutants defective in centromeric gene silencing yielded Sim3, an orthologue of the mammalian histone binding protein NASP (Dunleavy et al. 2007). Sim3 mutant yeast fail to load *Sp*CENP-A at centromeres, and it has been proposed to act as a chaperone in the pathway that delivers new *Sp*CENP-A to centromeres (Dunleavy et al. 2007).

1.3.2 De Novo *Centromere Formation*

Two very recent studies (Folco et al. 2008; Ishii et al. 2008) have addressed the issue of *de novo* centromere establishment in fission yeast. Delivery of naked DNA template containing centromere sequences (including *otr*, *imr*, and *cnt* sequences; see

Fig. 1.2) to wild-type fission yeast strains leads to acquisition of both kinetochore-forming chromatin (i.e., *Sp*CENP-A chromatin) and pericentromeric heterochromatin (Folco et al. 2008). In mutant strains, such as the ΔClr4 strain, that are unable to form pericentromeric heterochromatin, *Sp*CENP-A fails to assemble onto the naked DNA. If centromeres are initially formed in wild-type cells, however, the established *Sp*CENP-A-containing chromatin domain persists even after the removal of pericentromeric heterochromatin. *De novo* centromere assembly on existing chromatinized DNA along chromosome arms was addressed in a separate study where the entire centromere region of fission yeast chromosome 1 was deleted (Fig. 1.3; (Ishii et al. 2008)). Isolated survivors of centromere deletion were analyzed to determine the fate of the acentric chromosome 1. A subset of survivors recombined by telomere fusion with one of the other chromosomes (26%), while the major rescue pathway used in all other isolates was via neocentromere formation at sites lacking any centromere sequences. By contrast, rescue by neocentromere formation was rare (5–10% of survivors) in strains lacking the HP1 orthologue Swi6, Dicer (Dcr1), or the histone methyltransferase Clr4, each required for pericentromeric heterochromatin formation (Hall et al. 2002; Volpe et al. 2002). The findings of these two recent studies support the general notion that local chromatin environment is important for *de novo* centromere formation, but that once CENP-A marks the location of the centromere it is epigenetically maintained independently of pericentromeric heterochromatin.

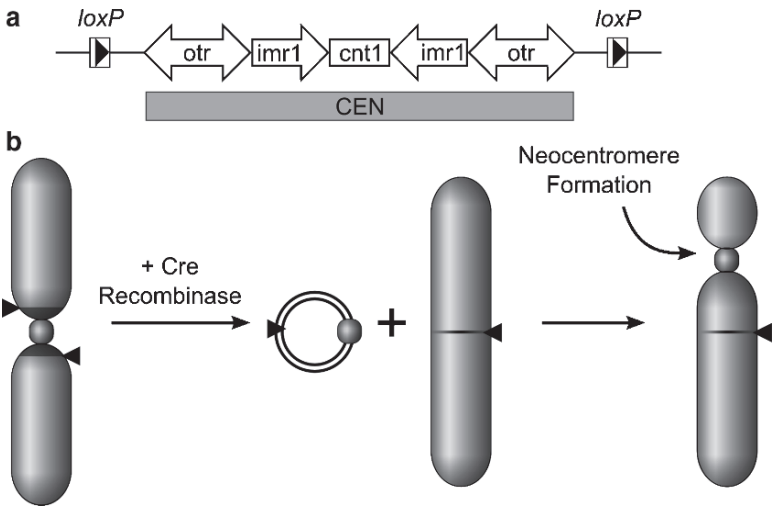


Fig. 1.3 Assay for neocentromere formation in *S. pombe*. (**a**) Engineered loxP sites flank the centromere of chromosome 1 for Cre-mediated excision (Ishii et al. 2008). (**b**) Inducible centromere excision leads to the formation of a minichromosome circle containing the centromere as well as an acentric chromosome. Cells surviving such centromere excision typically form a neocentromere on a chromosome arm site lacking any of the sequence elements found at the normal centromeres

1.4 The Maize Centromere

DNA repeats that underlie centromeres are greatly expanded (megabases) in plants and animals, as compared with yeast discussed earlier, and are typically comprised of more than one type of repeat. The centromeres of *Zea mays* are one such example with two sets of repeated elements that are found at the centromeres – the CentC repeats and the "maize centromere retroelements" (CRMs). CentC repeats are short (156 bp) and are repeated in tandem (Ananiev et al. 1998a). These repeats are intermingled with CRM repeats and form domains that range in size from ~0.3 to >2.8 Mbp (Jin et al. 2004; Chap. 6 in this book). In addition, the maize genome contains a wide variety of transposable elements, some of which are distributed uniformly throughout the arms and others are concentrated at centromeres (Ananiev et al. 1998a; Mroczek and Dawe 2003; Kato et al. 2004). All three classes of elements are found on each maize centromere, but their relative ratios vary widely (Ananiev et al. 1998a; Jin et al. 2004; Kato et al. 2004). The maize CENP-A orthologue, *Zm*CENP-A (also known as CenH3), is found, presumably incorporated into centromeric nucleosomes, on both CentC and CRM elements (Zhong et al. 2002; Jin et al. 2004). An additional constitutive centromere component, CENP-C, is also found associated with the same repetitive DNA sequences (Dawe et al. 1999). Intriguingly, CentC repeats and CRM element are transcribed, generating 40–200 nt small RNAs that are found stably bound to *Zm*CENP-A-containing chromatin (Topp et al. 2004). It should also be noted that centromeric repeats have been proposed to play a role in establishing RNAi dependent heterochromatin at rice centromeres (Neumann et al. 2007).

1.4.1 Epigenetic Centromere Silencing to Exit Breakage-Fusion-Bridge Cycles

Maize was recently used to assess epigenetic centromere inactivation (Han et al. 2006). The formation of dicentric chromosomes can result from a nondisjunction event, whereby two homologous chromosomes remain fused in meiosis. Such dicentric chromosomes subsequently enter the breakage-fusion-bridge cycle (BFB), which was originally described by McClintock (McClintock 1939, 1941). When the fusion occurs between an essential and a nonessential chromosome (called a "B chromosome" in maize), it is possible to study the consequences of such an event as its rearrangement does not have a deleterious phenotypic outcome (Zheng et al. 1999). The B chromosome derivative, B9-Dp9, forms a dicentric chromosome undergoing BFB cycles, but could potentially exit these cycles either by adding a telomere to one of the broken ends or by inactivating one of the two centromeres (Fig. 1.4). Out of the 23 chromosomes that have exited BFB cycles, six stable dicentric chromosomes were identified. Strikingly, all six had inactivated one of its centromeres (Han et al. 2006). These findings indicate that centromere inactivation through epigenetic silencing is prevalent even in the absence of any genetic selective pressure.

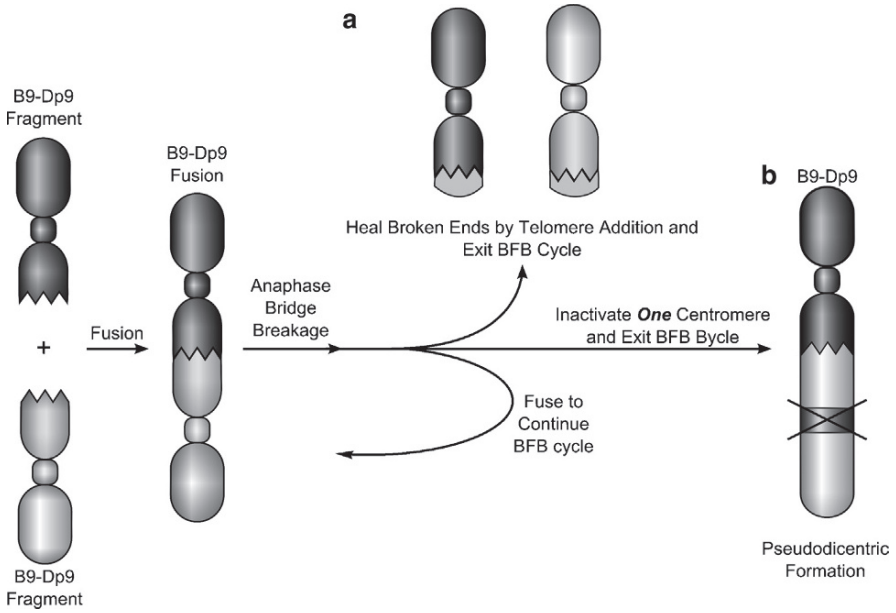


Fig. 1.4 Centromere inactivation in *Zea mays*. B9-Dp9 is a fusion of a centric fragment of the extra chromosome B9 and a region of chromosome 9 that contains an inverted duplication. Because of the presence of the inverted duplication, this chromosome derivative is prone to forming a dicentric chromosome and undergoing breakage-fusion-bridge (BFB) cycles. The dicentric chromosome may exit the BFB cycle either through (**a**) end healing prior to re-fusion or through (**b**) inactivation of one of the centromeres of a fused dicentric. In 6 of 23 exit events dicentric chromosomes remained and all 6 dicentrics had inactivated one of their centromeres (Han et al. 2006)

1.4.2 A Possible Role for DNA Methylation in Centromere Specification

DNA methylation of centromere sequences is well documented in plants (Hall et al. 2004). A recent study of maize and *Arabidopsis* centromeres showed that in each organism a subset of repetitive DNA is hypermethylated, but that the subset of centromeric DNA associated with CENP-A is hypomethylated (Zhang et al. 2008). While it is unclear whether or not these differences in DNA methylation play a role in centromere specification, the recent development of artificial chromosome technology in maize (Carlson et al. 2007) provides a potential system to assess genetic and epigenetic determinants of centromere establishment and maintenance in this plant.

1.4.3 Meiotic "Classical" Neocentromeres

Epigenetic chromosome segregation phenomena were studied in plants for decades before epigenetic centromere specification was known to occur in other

eukaryotic kingdoms. Interestingly, many of the original observations were made in maize. These include chromosome features that were the first to be termed "neocentromeres" (now termed "classical neocentromeres"; reviewed in (Dawe and Hiatt 2004)). As opposed to heritable neocentromeres that have been described in other kingdoms (and recently described in the barley plant as well (Nasuda et al. 2005)), classical neocentromeres are restricted to the plant kingdom. They lack known centromere components (including CENP-A and CENP-C orthologues (Dawe et al. 1999; Zhong et al. 2002)), are found in meiotic but *not* mitotic cells, lack the ability to mediate sister chromatid cohesion, and do not mediate chromosome biorientation on the meiotic spindle (Rhoades and Vilkomerson 1942; Yu et al. 1997; Hiatt et al. 2002; Mroczek et al. 2006). Classical neocentromeres of maize form on repetitive DNA sequences (so-called 180-bp repeats and TR1 elements) that are distinct in sequence from the CentC, CentA, and CRM that are typically found at bona fide maize centromeres (Peacock et al. 1981; Dennis and Peacock 1984; Ananiev et al. 1998b; Mroczek and Dawe 2003). When taken at face value, unlike neocentromeres in other systems, classical neocentromeres in plants are not directly relevant to the epigenetic pathways that specify the location of fully functional (and heritable) centromeres. It should be noted, however, that the sites used for classical neocentromere formation occur at large heterochromatic regions that are cytologically distinct from bulk chromatin, leading to their name "knobs." While the molecular mechanisms of classical neocentromeres remain unclear, these knobs are able to make microtubule attachments during female meiosis and move poleward during anaphase (Rhoades and Vilkomerson 1942; Yu et al. 1997). The attachments made at the knob neocentromere are unusual in that the attachment is made laterally instead of end-on, producing a thin chromatin fiber extension directed to the pole (Yu et al. 1997). Atypical meiotic spindle connections have also been reported in the holocentric centromere of the worm *Caenorhabditis elegans* at a cup-like kinetochore structure lacking underlying centromeric chromatin containing its CENP-A orthologue (Monen et al. 2005). It seems likely that atypical meiotic chromosome connections to the spindle are more pervasive than previously thought in the context of the enormous diversity found throughout eukaryotic chromosome biology.

1.5 The Fruit Fly Centromere

The identification of the DNA elements present in the *Drosophila melanogaster* centromere was made possible by utilizing a stable, nonessential X chromosome-derived minichromosome, *Dp1187*, which contains a functional centromere (Karpen and Spradling 1990, 1992; Tower et al. 1993; Le et al. 1995; Murphy and Karpen 1995; Sun et al. 1997, 2003). Using γ-irradiation induced breakage of *Dp1187*, its centromere was mapped to ~400 kb that contains both transposable elements as well as satellite repeats (Murphy and Karpen 1995; Sun et al. 1997, 2003). Generally the repeats are AT rich and are organized into discrete blocks where the smallest monomeric repeat unit is only 5 bp in length. Importantly, the repetitive sequence

elements mapped to the functional centromere are not limited to centromeric regions, thereby indicating that this DNA sequence is not sufficient for centromere specification in the fly. The transposable elements found at the *Dp1187* centromere are also found in several other regions of the genome, thus there is no single identifiable genetic element that is sufficient for centromere inheritance.

1.5.1 Spreading the Epigenetic Centromere Mark

With strong hints of an epigenetic mechanism at work at the fruit fly centromere, including the finding that centromere activity could be acquired on non-centromeric DNA (Williams et al. 1998), Karpen and colleagues developed a genetic system to assess centromere spreading onto DNA sequences previously lacking centromere function (Fig. 1.5; (Maggert and Karpen 2001)). Three chromosome derivatives with distinct test fragment positions were used: adjacent to centromeric chromatin, adjacent to pericentromeric heterochromatin, or adjacent to euchromatin. Upon the release of the test fragment from its chromosomal niche by irradiation, only the one located next to centromeric chromatin was able to maintain mitotic stability having acquired a functional centromere. This result indicates that the centromeric "mark" can spread along the chromosome. DNA modifications (such as DNA methylation), post-translational modification of histones or other components of centromeric chromatin could potentially generate this spreadable mark. One such centromeric chromatin component is *Dm*CENP-A (also referred to as CID), which has emerged as an attractive candidate to mark the fruit fly centromere (Henikoff et al. 2000; Blower and Karpen 2001).

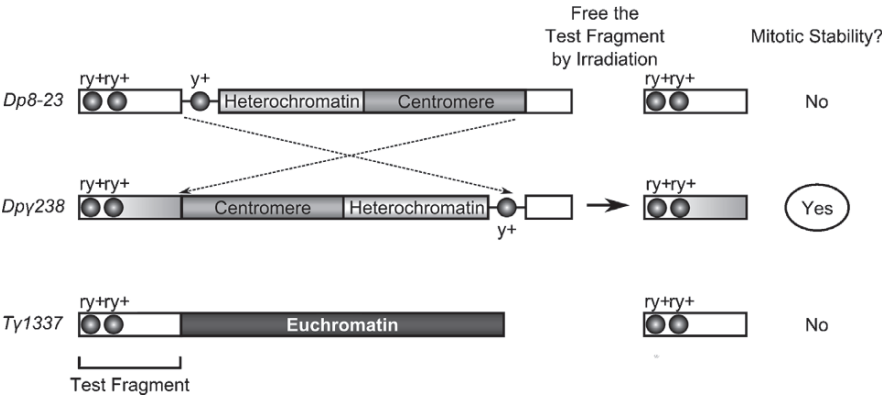


Fig. 1.5 Centromere spreading in *Drosophila melanogaster*. The position of the "test fragment" relative to the centromere on derivatives of the *Dp1187* minichromosome affects the ability of this fragment to obtain mitotic stability by acquiring centromere function (Maggert and Karpen 2001)

1.5.2 Higher-Order Chromatin Organization

The higher-order chromatin organization of the centromere in fruit flies has also been described and many aspects appear to be conserved in mammals. Consistent with prior chromosome stretching experiments that generated a repeating subunit model of higher order centromere structure (Zinkowski et al. 1991), the nucleosome arrangement on centromeres was found to contain *Dm*CENP-A-containing nucleosomes and H3-containing nucleosomes in interspersed blocks (Blower et al. 2002). This suggests a model where the chromatin at centromeres adopts a specialized three-dimensional conformation so that all of the *Dm*CENP-A-containing nucleosomes are clustered together on the surface of the centromere, at the foundation of the kinetochore, whereas the H3-containing nucleosomes are within the inner centromere, between sister kinetochores. Such organization has been envisioned to occur either by a looping or by a coiling organization (Blower et al. 2002). Characterization of the post-translational modification status of the intervening H3-containing nucleosomes revealed a de-enrichment for di- and tri- methylation of Lys9 on histone H3 relative to the enrichment at neighboring pericentromeric heterochromatin (Sullivan and Karpen 2004). These studies raise the central question of the modification state of centromeric nucleosomes. However, to date, the modification state of *Dm*CENP-A, as well as its relatives in other eukaryotes, remains largely elusive.

1.5.3 Centromere Marking by CENP-A-Containing Nucleosomes

Central questions remain unanswered regarding how the initial centromere mark is established. One prediction is that if an array of nucleosomes containing CENP-A epigenetically marks the centromere location, then the *de novo* formation of this array would be sufficient to establish a new centromere. To begin to address this prediction, *Dm*CENP-A was massively overexpressed and thereby forced to be incorporated into euchromatin in chromosome arms (Fig. 1.6; up to 70-fold over endogenous levels,

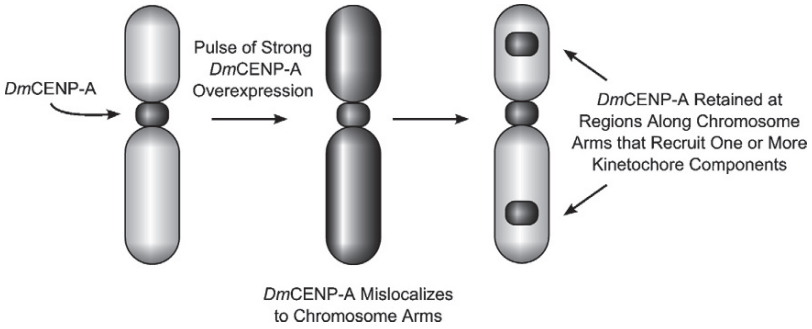


Fig. 1.6 Seeding new centromeric chromatin on chromosome arms. *Dm*CENP-A overexpression leads to misincorporation into chromosome arms, and these sites occasionally recruit one or more kinetochore components (Heun et al. 2006)

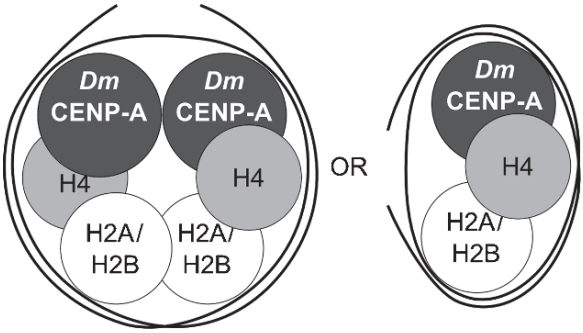


Fig. 1.7 Proposed composition of *Dm*CENP-A-containing nucleosomes

enough to replace the bulk of H3-containing nucleosomes present in euchromatin; (Heun et al. 2006)). Euchromatic CENP-A nucleosomes can be removed and subsequently degraded by the proteasome (Collins et al. 2004; Moreno-Moreno et al. 2006), leaving some chromosome arm sites that retained high levels of *Dm*CENP-A (Heun et al. 2006). While damage to endogenous centromeres was expected by diluting other endogenous centromere proteins during the initial spreading of *Dm*CENP-A-containing chromatin over the length of each chromosome, cells that survived contained chromosomes in which one or a few new regions enriched for *Dm*CENP-A persisted on chromosome arms. These ectopic sites occasionally recruited one or more kinetochore components. These findings support the hypothesis that establishing an array of CENP-A-containing nucleosomes generates the epigenetic mark that is sufficient for de novo centromere formation.

A recent study of *Dm*CENP-A-containing nucleosomes indicated that they are compact relative to their canonical counterparts containing H3 (Dalal et al. 2007). The available data are consistent with either a tetrameric nucleosome containing one copy of each histone or an octameric nucleosome of conventional histone stoichiometry that is converted into a more compact structure by the presence of *Dm*CENP-A (Fig. 1.7; (Black and Bassett 2008; Dalal et al. 2007)). In either case, the unique physical properties conferred by *Dm*CENP-A, which distinguish centromeric nucleosomes from bulk chromatin, are central to its ability to epigenetically mark the fruit fly centromere.

1.6 The Human Centromere

Functional human centromeres are typically found in regions containing megabase stretches of a specific form of repetitive DNA, termed α-satellite, where the smallest monomer repeat unit is 171 bp (for a review on α-satellite DNA, see Willard (1991). Naturally occurring rearrangements of the human X chromosome proximal to the functional centromere revealed that other types of DNA satellite sequences surrounding its α-satellite domain can be removed, but the α-satellite domain is retained and

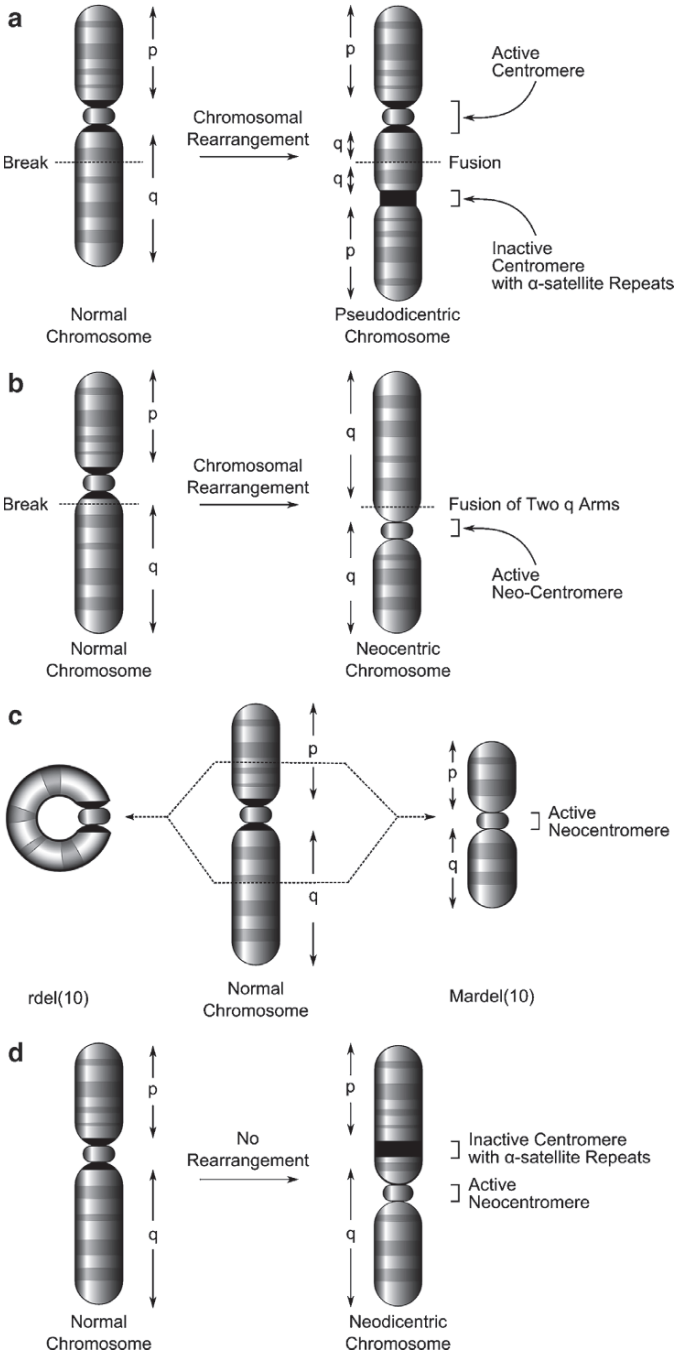
contains centromeric proteins (Schueler et al. 2001). A potential connection between centromeric DNA sequences and other centromeric proteins is CENP-B, which is a sequence-specific DNA binding protein that recognizes a 17 bp sequence, termed the CENP-B box, that is found in α-satellite repeats (Earnshaw and Rothfield 1985; Valdivia and Brinkley 1985; Earnshaw et al. 1987; Masumoto et al. 1989). However, it is unclear which, if any, other centromere components are recruited by CENP-B. Furthermore, functional human centromeres are defined not by α-satellite sequences or CENP-B, but by the presence of other constitutive protein components (such as CENP-A and CENP-C) and their ability to build a kinetochore at mitosis. As is the case for all centromeres that have been studied in other eukaryotes, with the exception of budding yeast, there are important genetic and epigenetic components to be considered in any discussion of centromere identity. In the study of human centromeres, multiple avenues of investigation have been fruitful: understanding centromere silencing and de novo centromere formation in patients with abnormal chromosomes, testing the requirements for forming human artificial chromosomes (HACs), physical characterization of the nucleosomes that are the building blocks of centromeric chromatin, and the elucidation of the cellular pathway that maintains centromere identity. In this section we discuss many of the major advances in each of these areas.

1.6.1 Chromosomal Rearrangements

Chromosomal abnormalities found in the human population arise from diverse forms of alterations, including duplications, inversions, deletions, and translocations. The resulting chromosome products may lack the centromere or contain more than one centromere. Both cases present a major problem for chromosome segregation at cell division, and there are clear examples where centromere activity is silenced or generated *de novo* to ensure that one and only one active centromere exists on the abnormal chromosome.

When a dicentric chromosome arises via a fusion event where the two centromere loci are spaced sufficiently far apart (>12 Mbps; (Sullivan and Willard 1998)), one of the two centromeres is inactivated to avoid multiple attachments to the spindle that would have a propensity to cause chromosome breakage on the spindle (akin to the behavior of the dicentrics in the breakage-fusion bridge cycles in maize, described in Sect. 1.4.1). Centromere inactivation, generating a pseudodicentric chromosome with one functional centromere, does not require additional DNA rearrangements at the centromere locus. Rather, megabase stretches of α-satellite sequences remain at the inactive locus, suggesting an epigenetic mechanism of inactivation that warrants further investigation (Fig. 1.8a).

The reciprocal chromosome segregation problem arises in the case of acentric chromosome fragments lacking an endogenous centromere (Fig. 1.8b). Such segments would be genetically unstable unless they are able to rapidly generate a new functional centromere. The first descriptions of such neocentromeres emerged in the 1990s and there are now >90 known cases of human neocentromeres, with representative


a
p
Break
q
Normal
Chromosome
Chromosomal
Rearrangement
p
q
q
p
Active
Centromere
Fusion
Inactive
Centromere
with α-satellite Repeats
Pseudodicentric
Chromosome
b
p
Break
q
Normal
Chromosome
Chromosomal
Rearrangement
q
q
Fusion of Two q Arms
Active
Neo-Centromere
Neocentric
Chromosome
c
rdel(10)
p
q
Normal
Chromosome
p
q
Active
Neocentromere
Mardel(10)
d
p
q
Normal
Chromosome
No
Rearrangement
p
q
Inactive Centromere
with α-satellite Repeats
Active
Neocentromere
Neodicentric
Chromosome

cases on nearly every chromosome (Marshall et al. 2008; Chap. 5 in this book). The best characterized neocentromere is on a chromosome 10 fragment, termed mardel(10) (Fig. 1.8c; (Voullaire et al. 1993; du Sart et al. 1997; Barry et al. 1999, 2000; Chueh et al. 2005; Lo et al. 2001a)). Mardel(10) is mitotically stable (Voullaire et al. 1993) and lacks detectable α-satellite DNA. The other cases where the neocentromere has been closely mapped (Lo et al. 2001b; Alonso et al. 2003; Cardone et al. 2006) support the general view that neocentromere formation does not require any further chromosomal rearrangements that would yield new locations of α-satellite DNA or any other detectable repetitive sequences. Rather, the prevailing view is that neocentromere formation in humans occurs by an epigenetic mechanism.

Of the proteins that discretely localize to centromeres, CENP-B is unique in that it follows α-satellite DNA sequences irrespective of the functional state of the centromere. In other words, CENP-B remains at the silenced centromeres of pseudodicentric chromosomes (Earnshaw et al. 1989; Sullivan and Schwartz 1995; Warburton et al. 1997) and is not recruited to neocentromeres (Voullaire et al. 1993; Saffery et al. 2000). Along with the finding that the mouse version of CENP-B is dispensable for viability as well as meiotic and mitotic centromere function, a general view has emerged that CENP-B and its recognition element (the CENP-B box) within α-satellite repeats is irrelevant to centromere function. This view has been challenged by experiments with artificial chromosomes to monitor the establishment of centromere identity (discussed below in Sect. 1.6.3). Other proteins, such as CENP-A, CENP-C, and CENP-H, that discretely and constitutively localize to normal centromeres track with functional centromeres on rearranged chromosomes: absent from inactive centromeres and present at neocentromeres (Sullivan and Schwartz 1995; Warburton et al. 1997; Sugata et al. 2000; Warburton et al. 2000).

1.6.2 Neodicentric Chromosomes

Both the silencing of a centromere in dicentric chromosomes, as well as the formation of neocentromeres in acentric chromosomes in human patients is expected to occur under strong selective pressure, as each of these epigenetic events rescues the impacted chromosome from peril at cell division. However, more recent findings with intact

Fig. 1.8 Pseudodicentric, neocentromeric, and neodicentric chromosomes. (**a**) Dicentric chromosomes typically arise through chromosome fusion. When this happens, the dicentric chromosome may achieve mitotic stability and avoid breakage on the spindle by inactivating one of its centromeres. This forms a pseudodicentric chromosome that contains two distinct α-satellite loci (*shaded in black*), but only one of which acts as a functional centromere. (**b**) Genetic rearrangement leading to the formation of an acentric chromosome. Mitotic stability is regained through neocentromere formation at a locus lacking α-satellite repeats. (**c**) The Mardel(10) neocentromeric chromosome (*right*) was the acentric product of an internal recombination event that looped out the endogenous centromere (circular mini-chromosome, rdel(10), *left*). (**d**) Epigenetic centromere repositioning on neodicentric chromosomes occurs when the functional centromere relocates to a non-alphoid locus in the absence of any DNA rearrangements

chromosomes where centromere location has repositioned to a chromosome arm site lacking α-satellite sequences – so-called *pseudodicentric/neocentromeric* or *neodicentric* chromosomes – have called this notion into question. Following three suggestive examples on the Y chromosome where the centromere had relocated (Bukvic et al. 1996; Rivera et al. 1996; Tyler-Smith et al. 1999), two recent descriptions of autosomes with such centromere repositioning events were described (Amor et al. 2004; Ventura et al. 2004). One of these involved a repositioned centromere on chromosome 4, where the original centromere location, now epigenetically silenced, retains >1 Mbp of α-satellite DNA (Fig. 1.8d; (Amor et al. 2004)). The mechanism of such centromere repositioning is unclear, but any simple scenarios not requiring chromosomal fragment intermediates would lead to a model where either centromere inactivation or neocentromere formation can occur without any selective pressure. Beyond this, the fact that reversion of the centromere back to the original location does not occur in the individuals, their offspring, or even after long term culturing of their cells (Amor et al. 2004; Ventura et al. 2004) lends strong support to a model of centromere identity wherein once DNA is marked by an array of CENP-A-containing nucleosomes, a new centromere location is epigenetically maintained in perpetuity.

1.6.3 Artificial Chromosomes

Human artificial chromosomes (HACs) hold the promise as vectors for gene delivery, and provide powerful tools for fundamental investigations of centromere structure and function. Acquisition of a functional centromere appears to be the most important step in generating a functional HAC, as the HAC templates that fail to achieve autonomous chromosome segregation behavior integrate into an existing chromosome (Haaf et al. 1992; Larin et al. 1994; Warburton and Cooke 1997). While the efficiency of HAC formation is low even in the best cases using α-satellite-containing templates for centromere formation, non-alphoid templates completely fail to form HACs (Harrington et al. 1997; Henning et al. 1999; Ikeno et al. 1998; Masumoto et al. 1998; Ebersole et al. 2000; Saffery et al. 2001; Schueler et al. 2001; Grimes et al. 2002). Starting with a heroic cloning effort, Masumoto and colleagues mutated the CENP-B box of a single α-satellite monomer and multimerized it to generate an ~70 kb HAC template mimicking α-satellite DNA but lacking any functional CENP-B boxes (Ohzeki et al. 2002). These mutant HAC templates fail to form functional HACs, as do α-satellite HAC templates in cells lacking endogenous CENP-B protein (Ohzeki et al. 2002). HAC formation remains a rare event, with most integrating into host chromosomes, as mentioned earlier. A clue as to why this might be so was revealed by monitoring the recruitment of CENP-B and CENP-A to HAC templates by chromatin immunoprecipitation in the days following HAC template transfection (Fig. 1.9; (Okada et al. 2007)). While full levels of CENP-B are recruited in the first timepoint (one day following transfection), CENP-A is not recruited appreciably for four days, suggesting that several cell divisions are required to establish centromeric chromatin in mammalian cells.

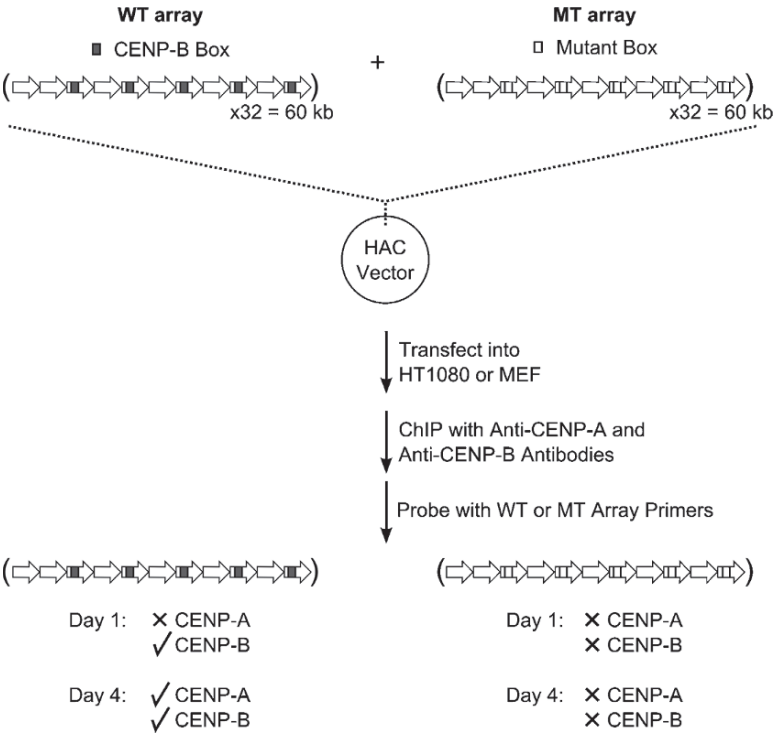


Fig. 1.9 CENP-B is involved in an early step in centromere establishment on HAC DNA. CENP-B rapidly accumulates on wild type (WT) but not engineered CENP-B-box mutant (MT) arrays in a contiguous HAC construct. Initial CENP-A assembly does not occur for several days, and it is only able to assemble on WT arrays bound by CENP-B (Okada et al. 2007)

Once established, centromeric chromatin in mammalian cells is sensitive to specific forms of perturbation in neighboring chromatin domains (Fig. 1.10; (Nakano et al. 2008)). HACs engineered with interspersed α-satellite repeats containing either CENP-B boxes or tetracycline operator sites (tetO) enable the targeting of proteins of interest fused to the tetracycline repressor (tetR). While targeting of a transcriptional activator (tTA) had a modest affect on HAC stability, a dramatic loss of HAC stability was observed by targeting of the transcriptional silencing domain from the Kid1 protein. This correlated with a loss of CENP-A, CENP-B, and CENP-C, a local accumulation of H3 nucleosomes modified with a dimethylation at Lys4, and a spreading of the silenced chromatin into the neighboring antibiotic resistance gene on the HAC construct. Direct targeting of heterochromatin protein 1-α (HP1α) causes a similar loss of CENP-C, supporting the notion that HP1α accumulation commonly found in pericentromere regions is mutually exclusive from the kinetochore-forming portion of the centromere responsible for its specification.

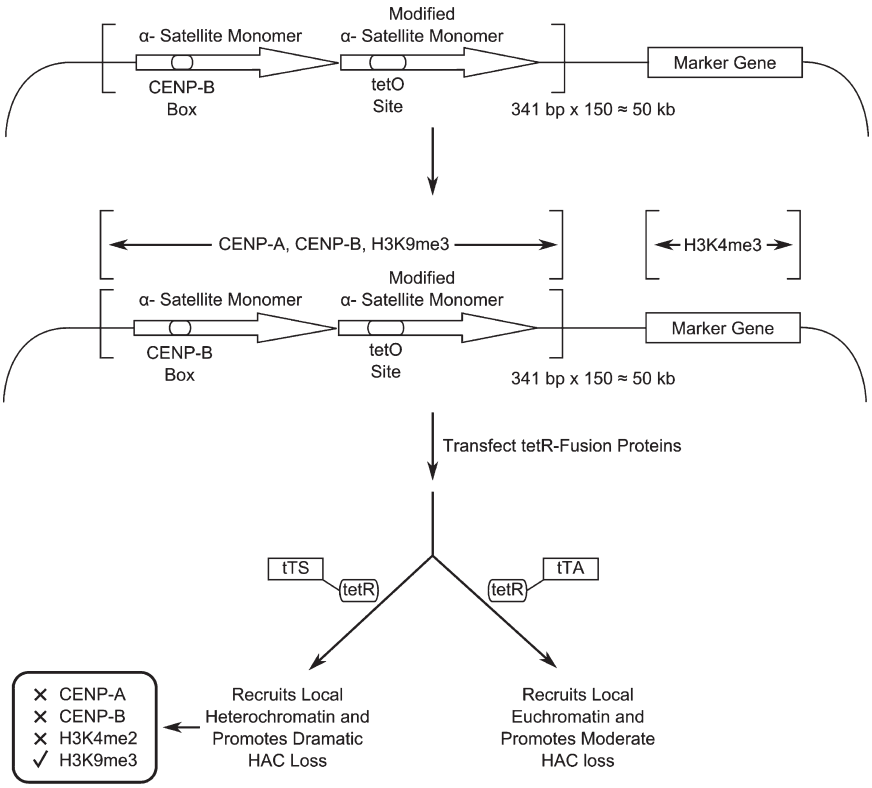


Fig. 1.10 Inactivating centromeres on engineered HACs. Tetracycline repressor (tet-R) fusions are used to target transcriptional activators (tTA) or silencers (tTS) to functional, mitotically stable HACs carrying tetracycline operator (tet-O). Transcriptional silencers promote heterochromatin formation, which destabilizes the HAC (Nakano et al. 2008)

1.6.4 Mechanisms to Maintain Centromere Identity

CENP-A has emerged as the key determinant of centromere identity as it is always found at functional centromeres, is absent from inactive centromere, and is a subunit of an octameric histone core that wraps DNA. Critical questions have been pursued in recent years. How does CENP-A physically differentiate the chromatin into which it is assembled from the rest of the chromosome? How is newly expressed CENP-A protein targeted to centromeres? When during the cell cycle does this occur?

To serve as an epigenetic determinant of centromere identity, CENP-A must distinguish the chromatin into which it is assembled from bulk chromatin at the level of an individual nucleosome and/or at the level of the array of 10^3–10^4 nucleosomes it forms at each centromere. This could be achieved in a manner similar to other well-studied epigenetic marks, such as those carried by self-perpetuating histone

modifications (Hake and Allis 2006). These post-translational modification-based marks are recognized by specific chromatin binding proteins that drive the local recruitment of the histone modifying enzymes themselves, thus perpetuating the epigenetic mark. CENP-A may mark the centromere in a related but physically distinct manner, taking advantage of the unique conformational rigidity of the nucleosomes into which it is assembled (Black et al. 2007a). This rigidity is conferred by the loop 1 and α2-helix within its histone fold domain, a region termed the CENP-A targeting domain (CATD) that includes 22 amino acid changes relative to histone H3 (Fig. 1.11; (Black et al. 2004, 2007a)). If this unique structure recruits a protein(s) that in turn promotes the recruitment of newly expressed CENP-A, then the important unit of the epigenetic centromere mark is an individual nucleosome. If, however, the unique structure drives self–self interactions that culminate in higher order chromatin folding (such as the coalesced CENP-A array proposed by Sullivan, Karpen and colleagues; (Blower et al. 2002; Schueler and Sullivan 2006)) that is recognized by proteins participating in CENP-A recruitment, then the important unit of the epigenetic centromere mark is the higher order CENP-A nucleosome array. In either scenario, components of the CENP-A nucleosome associated complex (CENP-A^{NAC}) (Fig. 1.12; CENP-C, CENP-H, CENP-M, CENP-N, CENP-T and

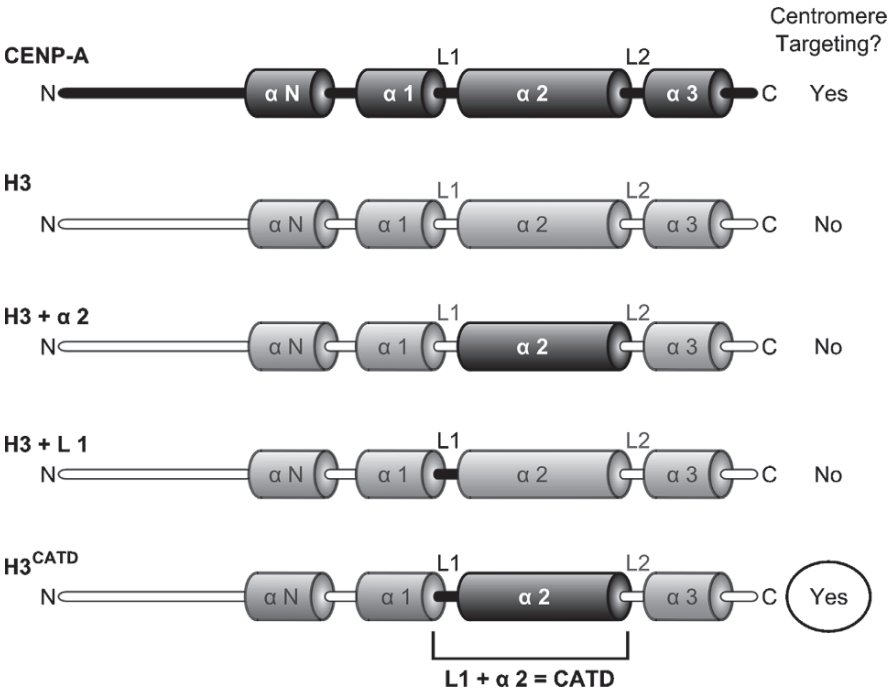


Fig. 1.11 Essential structural elements of the CENP-A targeting domain (CATD). Loop 1 (L1) and the α2 helix of human CENP-A together form the CATD that is sufficient to direct H3 to the centromere (Black et al. 2004; Black et al. 2007b)

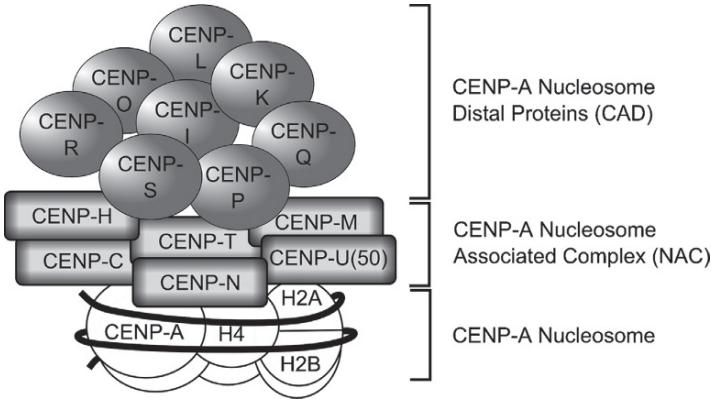


Fig. 1.12 The CENP-A nucleosome associated complexes. CENP-A nucleosomes co-purify with members of the CENP-A^{NAC} that are constitutively found at centromeres (Foltz et al. 2006; Obuse et al. 2004; Okada et al. 2006). A more distal complex, CENP-A^{CAD}, contains several additional constitutive centromere components

CENP-U(50) (Obuse et al. 2004; Foltz et al. 2006; Okada et al. 2006)) are excellent candidate molecules for recognizing the mark specified by the unique chromatin generated by the incorporation of CENP-A.

The *cis*-acting information for targeting newly expressed CENP-A to functional centromeres is contained within the CATD (Black et al. 2004, 2007b). Since specific cellular pathways exist for regulating the deposition of the bulk H3 variants H3.1 and H3.3, it seems likely that a mechanism exists to recognize CENP-A via the CATD in a pathway that maintains the epigenetic centromere mark. H3.1 and H3.3 are nearly identical, varying at five amino acid positions, yet they are recognized by different histone chaperone complexes: CAF1 and HIRA, respectively (Smith and Stillman 1989; Ray-Gallet et al. 2002; Tagami et al. 2004). CAF1 loading of H3.1 is coupled to replication, while HIRA loading of H3.3 occurs throughout the cell cycle (Worcel et al. 1978; Wu et al. 1982; Ahmad and Henikoff 2002; Ray-Gallet et al. 2002; Tagami et al. 2004). The loading of newly expressed CENP-A is uncoupled from DNA replication (Shelby et al. 2000). Rather, it is produced early in the G2 phase of the cell cycle (Shelby et al. 2000) but does not load onto centromeres until late telophase of mitosis and the first few hours of the subsequent G1 phase (Jansen et al. 2007; Schuh et al. 2007; Hemmerich et al. 2008). At a minimum, it is expected that the CATD delineates newly expressed CENP-A from the H3.1 and H3.3 chromatin deposition pathways. It is also quite likely that the CATD accesses a dedicated centromeric chromatin assembly pathway. At the final step of the pathway, assembly into centromeric nucleosomes, a centromere priming event has been proposed to involve the human orthologue of the *S. pombe* Mis18 protein and Mis18BP1/KNL2 (Fujita et al. 2007; Maddox et al. 2007), each of which is required for new CENP-A nucleosome assembly, and each of which transiently visit the centromere during a time window overlapping of CENP-A assembly.

1.7 Outlook

Our knowledge of how centromeres are specified has come from the study of many diverse eukaryotic species. We have highlighted five species that have been particularly helpful in shaping the current view. Certainly, many key questions remain unanswered. For example, the fundamental unit of centromeric chromatin, the CENP-A nucleosome, remains to be structurally elucidated on the atomic level, and its very composition has emerged recently as an area that requires additional experimentation. Furthermore, despite recent progress, the pathway for CENP-A assembly into nucleosomes is not well understood in any eukaryote. The paradox of the centromere persists due to the seemingly discordant findings that α-satellite sequences are dispensable for centromere function on naturally occurring chromosome variants, yet the same sequences are required for detectable levels of de novo HAC formation. The physical relationship between CENP-A-containing nucleosomes and α-satellite DNA, therefore, requires further investigation. All these questions are fundamental to our understanding of the chromosomal locus that ensures the integrity of the genome at cell division.

Note added in proof We call attention to two studies (Foltz et al. 2009, Centromere specific assembly of CENP-A nucleosomes is mediated by HJURP, Cell, in press; Dunleavy et al., 2009, HJURP, a key CENP-A-partner for maintenance and deposition of CENP-A at centromeres at late telophase/G1, Cell, in press) embering during the editing and production of this chapter. These studies independently identified HJURP as a trans-acting histone chaperone that is essential for CENP-A deposition at human centromeres. Furthermore, Foltz and colleagues found that recognition of CENP-A by HJURP is mediated through the CATD.

Acknowledgements Work in the Black Laboratory is supported by a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund and a grant (GM82989) from the NIH.

References

Ahmad K, Henikoff S (2002) The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell 9:1191–1200

Allshire RC, Javerzat JP, Redhead NJ, Cranston G (1994) Position effect variegation at fission yeast centromeres. Cell 76:157–169

Allshire RC, Nimmo ER, Ekwall K, Javerzat JP, Cranston G (1995) Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. Genes Dev 9:218–233

Alonso A, Mahmood R, Li S, Cheung F, Yoda K, Warburton PE (2003) Genomic microarray analysis reveals distinct locations for the CENP-A binding domains in three human chromosome 13q32 neocentromeres. Hum Mol Genet 12:2711–2721

Amor DJ, Bentley K, Ryan J, Perry J, Wong L, Slater H, Choo KH (2004) Human centromere repositioning "in progress". Proc Natl Acad Sci USA 101:6542–6547

Ananiev EV, Phillips RL, Rines HW (1998a) Chromosome-specific molecular organization of maize (*Zea mays L.*) centromeric regions. Proc Natl Acad Sci USA 95:13073–13078

Ananiev EV, Phillips RL, Rines HW (1998b) A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons. Proc Natl Acad Sci USA 95:10785–10790

Baker RE, Fitzgerald-Hayes M, O'Brien TC (1989) Purification of the yeast centromere binding protein CP1 and a mutational analysis of its binding site. J Biol Chem 264:10843–10850

Barry AE, Howman EV, Cancilla MR, Saffery R, Choo KH (1999) Sequence analysis of an 80 kb human neocentromere. Hum Mol Genet 8:217–227

Barry AE, Bateman M, Howman EV, Cancilla MR, Tainton KM, Irvine DV, Saffery R, Choo KH (2000) The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification. Genome Res 10:832–838

Black BE, Bassett EA (2008) The histone variant CENP-A and centromere specification. Curr Opin Cell Biol 20:91–100

Black BE, Foltz DR, Chakravarthy S, Luger K, Woods VL Jr, Cleveland DW (2004) Structural determinants for generating centromeric chromatin. Nature 430:578–582

Black BE, Brock MA, Bedard S, Woods VL Jr, Cleveland DW (2007a) An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. Proc Natl Acad Sci USA 104:5008–5013

Black BE, Jansen LE, Maddox PS, Foltz DR, Desai AB, Shah JV, Cleveland DW (2007b) Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol Cell 25:309–322

Bloom KS, Carbon J (1982) Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. Cell 29:305–317

Blower MD, Karpen GH (2001) The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat Cell Biol 3:730–739

Blower MD, Sullivan BA, Karpen GH (2002) Conserved organization of centromeric chromatin in flies and humans. Dev Cell 2:319–330

Bram RJ, Kornberg RD (1987) Isolation of a *Saccharomyces cerevisiae* centromere DNA-binding protein, its human homolog, and its possible role as a transcription factor. Mol Cell Biol 7:403–409

Bukvic N, Susca F, Gentile M, Tangari E, Iannirubertо A, Guanti G (1996) An unusual dicentric Y chromosome with a functional centromere with no detectable alpha-satellite. Hum Genet 97:453–456

Cai MJ, Davis RW (1989) Purification of a yeast centromere-binding protein that is able to distinguish single base-pair mutations in its recognition site. Mol Cell Biol 9:2544–2550

Camahort R, Li B, Florens L, Swanson SK, Washburn MP, Gerton JL (2007) Scm3 is essential to recruit the histone h3 variant cse4 to centromeres and to maintain a functional kinetochore. Mol Cell 26:853–865

Carbon J, Clarke L (1984) Structural and functional analysis of a yeast centromere (CEN3). J Cell Sci Suppl 1:43–58

Cardone MF, Alonso A, Pazienza M, Ventura M, Montemurro G, Carbone L, de Jong PJ, Stanyon R, D'Addabbo P, Archidiacono N, She X, Eichler EE, Warburton PE, Rocchi M (2006) Independent centromere formation in a capricious, gene-free domain of chromosome 13q21 in Old World monkeys and pigs. Genome Biol 7:R91

Carlson SR, Rudgers GW, Zieler H, Mach JM, Luo S, Grunden E, Krol C, Copenhaver GP, Preuss D (2007) Meiotic transmission of an in vitro-assembled autonomous maize minichromosome. PLoS Genet 3:1965–1974

Castillo AG, Mellone BG, Partridge JF, Richardson W, Hamilton GL, Allshire RC, Pidoux AL (2007) Plasticity of fission yeast CENP-A chromatin driven by relative levels of histone H3 and H4. PLoS Genet 3:e121

Chen ES, Saitoh S, Yanagida M, Takahashi K (2003) A cell cycle-regulated GATA factor promotes centromeric localization of CENP-A in fission yeast. Mol Cell 11:175–187

Chikashige Y, Kinoshita N, Nakaseko Y, Matsumoto T, Murakami S, Niwa O, Yanagida M (1989) Composite motifs and repeat symmetry in *S. pombe* centromeres: direct analysis by integration of NotI restriction sites. Cell 57:739–751

Chueh AC, Wong LH, Wong N, Choo KH (2005) Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. Hum Mol Genet 14:85–93

Clarke L, Baum MP (1990) Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences. Mol Cell Biol 10:1863–1872

Clarke L, Carbon J (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature 287:504–509

Clarke L, Carbon J (1983) Genomic substitutions of centromeres in *Saccharomyces cerevisiae*. Nature 305:23–28

Clarke L, Amstutz H, Fishel B, Carbon J (1986) Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. Proc Natl Acad Sci USA 83:8253–8257

Clarke L, Baum M, Marschall LG, Ngan VK, Steiner NC (1993) Structure and function of *Schizosaccharomyces pombe* centromeres. Cold Spring Harb Symp Quant Biol 58:687–695

Collins KA, Furuyama S, Biggins S (2004) Proteolysis contributes to the exclusive centromere localization of the yeast Cse4/CENP-A histone H3 variant. Curr Biol 14:1968–1972

Cottarel G, Shero JH, Hieter P, Hegemann JH (1989) A 125-base-pair CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae*. Mol Cell Biol 9:3342–3349

Cumberledge S, Carbon J (1987) Mutational analysis of meiotic and mitotic centromere function in *Saccharomyces cerevisiae*. Genetics 117:203–212

Dalal Y, Wang H, Lindsay S, Henikoff S (2007) Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. PLoS Biol 5:e218

Dawe RK, Hiatt EN (2004) Plant neocentromeres: fast, focused, and driven. Chromosome Res 12:655–669

Dawe RK, Reed LM, Yu HG, Muszynski MG, Hiatt EN (1999) A maize homolog of mammalian CENPC is a constitutive component of the inner kinetochore. Plant Cell 11:1227–1238

Dennis ES, Peacock WJ (1984) Knob heterochromatin homology in maize and its relatives. J Mol Evol 20:341–350

du Sart D, Cancilla MR, Earle E, Mao JI, Saffery R, Tainton KM, Kalitsis P, Martyn J, Barry AE, Choo KH (1997) A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. Nat Genet 16:144–153

Dunleavy EM, Pidoux AL, Monet M, Bonilla C, Richardson W, Hamilton GL, Ekwall K, McLaughlin PJ, Allshire RC (2007) A NASP (N1/N2)-related protein, Sim3, binds CENP-A and is required for its deposition at fission yeast centromeres. Mol Cell 28:1029–1044

Earnshaw W, Bordwell B, Marino C, Rothfield N (1986) Three human chromosomal autoantigens are recognized by sera from patients with anti-centromere antibodies. J Clin Invest 77: 426–430

Earnshaw WC, Cooke CA (1989) Proteins of the inner and outer centromere of mitotic chromosomes. Genome 31:541–552

Earnshaw WC, Rothfield N (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma 91:313–321

Earnshaw WC, Sullivan KF, Machlin PS, Cooke CA, Kaiser DA, Pollard TD, Rothfield NF, Cleveland DW (1987) Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. J Cell Biol 104:817–829

Earnshaw WC, Ratrie H, 3rd, Stetten G (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. Chromosoma 98:1–12

Ebersole TA, Ross A, Clark E, McGill N, Schindelhauer D, Cooke H, Grimes B (2000) Mammalian artificial chromosome formation from circular alphoid input DNA does not require telomere repeats. Hum Mol Genet 9:1623–1631

Eichler EE (1999) Repetitive conundrums of centromere structure and function. Hum Mol Genet 8:151–155

Ekwall K, Olsson T, Turner BM, Cranston G, Allshire RC (1997) Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. Cell 91:1021–1032

Fishel B, Amstutz H, Baum M, Carbon J, Clarke L (1988) Structural organization and functional analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. Mol Cell Biol 8:754–763

Fitzgerald-Hayes M (1987) Yeast centromeres. Yeast 3:187–200

Fitzgerald-Hayes M, Buhler JM, Cooper TG, Carbon J (1982a) Isolation and subcloning analysis of functional centromere DNA (CEN11) from Saccharomyces cerevisiae chromosome XI. Mol Cell Biol 2:82–87

Fitzgerald-Hayes M, Clarke L, Carbon J (1982b) Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. Cell 29:235–244

Folco HD, Pidoux AL, Urano T, Allshire RC (2008) Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. Science 319:94–97

Foltz DR, Jansen LE, Black BE, Bailey AO, Yates JR III, Cleveland DW (2006) The human CENP-A centromeric nucleosome-associated complex. Nat Cell Biol 8:458–469

Fujita Y, Hayashi T, Kiyomitsu T, Toyoda Y, Kokubu A, Obuse C, Yanagida M (2007) Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. Dev Cell 12:17–30

Funk M, Hegemann JH, Philippsen P (1989) Chromatin digestion with restriction endonucleases reveals 150-160 bp of protected DNA in the centromere of chromosome XIV in Saccharomyces cerevisiae. Mol Gen Genet 219:153–160

Furuyama S, Biggins S (2007) Centromere identity is specified by a single centromeric nucleosome in budding yeast. Proc Natl Acad Sci USA 104:14706–14711

Gaudet A, Fitzgerald-Hayes M (1987) Alterations in the adenine-plus-thymine-rich region of CEN3 affect centromere function in *Saccharomyces cerevisiae*. Mol Cell Biol 7:68–75

Goh PY, Kilmartin JV (1993) NDC10: a gene involved in chromosome segregation in *Saccharomyces cerevisiae*. J Cell Biol 121:503–512

Grimes BR, Rhoades AA, Willard HF (2002) Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. Mol Ther 5:798–805

Haaf T, Warburton PE, Willard HF (1992) Integration of human alpha-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. Cell 70:681–696

Hahnenberger KM, Baum MP, Polizzi CM, Carbon J, Clarke L (1989) Construction of functional artificial minichromosomes in the fission yeast *Schizosaccharomyces pombe*. Proc Natl Acad Sci USA 86:577-581

Hahnenberger KM, Carbon J, Clarke L (1991) Identification of DNA regions required for mitotic and meiotic functions within the centromere of *Schizosaccharomyces pombe* chromosome I. Mol Cell Biol 11:2206–2215

Hajra S, Ghosh SK, Jayaram M (2006) The centromere-specific histone variant Cse4p (CENP-A) is essential for functional chromatin architecture at the yeast 2-microm circle partitioning locus and promotes equal plasmid segregation. J Cell Biol 174:779–790

Hake SB, Allis CD (2006) Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proc Natl Acad Sci USA 103:6428–6435

Hall AE, Keith KC, Hall SE, Copenhaver GP, Preuss D (2004) The rapidly evolving field of plant centromeres. Curr Opin Plant Biol 7:108–114

Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI (2002) Establishment and maintenance of a heterochromatin domain. Science 297:2232–2237

Han F, Lamb JC, Birchler JA (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. Proc Natl Acad Sci USA 103:3238–3243

Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF (1997) Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. Nat Genet 15:345–355

Hayashi T, Fujita Y, Iwasaki O, Adachi Y, Takahashi K, Yanagida M (2004) Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell 118:715–729

Hemmerich P, Weidtkamp-Peters S, Hoischen C, Schmiedeberg L, Erliandri I, Diekmann S (2008) Dynamics of inner kinetochore assembly and maintenance in living cells. J Cell Biol 180:1101–1114

Henikoff S, Ahmad K, Platero JS, van Steensel B (2000) Heterochromatic deposition of centromeric histone H3-like proteins. Proc Natl Acad Sci USA 97:716–721

Henning KA, Novotny EA, Compton ST, Guan XY, Liu PP, Ashlock MA (1999) Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. Proc Natl Acad Sci USA 96:592–597

Heun P, Erhardt S, Blower MD, Weiss S, Skora AD, Karpen GH (2006) Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Dev Cell 10:303–315

Hiatt EN, Kentner EK, Dawe RK (2002) Independently regulated neocentromere activity of two classes of tandem repeat arrays. Plant Cell 14:407–420

Hieter P, Pridmore D, Hegemann JH, Thomas M, Davis RW, Philippsen P (1985) Functional selection and analysis of yeast centromeric DNA. Cell 42:913–921

Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H (1998) Construction of YAC-based mammalian artificial chromosomes. Nat Biotechnol 16:431–439

Ishii K, Ogiyama Y, Chikashige Y, Soejima S, Masuda F, Kakuma T, Hiraoka Y, Takahashi K (2008) Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. Science 321:1088–1091

Jansen LE, Black BE, Foltz DR, Cleveland DW (2007) Propagation of centromeric chromatin requires exit from mitosis. J Cell Biol 176:795–805

Jayaram M, Li YY, Broach JR (1983) The yeast plasmid 2mu circle encodes components required for its high copy propagation. Cell 34:95–104

Jayaram M, Sutton A, Broach JR (1985) Properties of REP3: a cis-acting locus required for stable propagation of the *Saccharomyces cerevisiae* plasmid 2 microns circle. Mol Cell Biol 5:2466–2475

Jehn B, Niedenthal R, Hegemann JH (1991) In vivo analysis of the Saccharomyces cerevisiae centromere CDEIII sequence: requirements for mitotic chromosome segregation. Mol Cell Biol 11:5212–5221

Jiang WD, Philippsen P (1989) Purification of a protein binding to the CDEI subregion of *Saccharomyces cerevisiae* centromere DNA. Mol Cell Biol 9:5585–5593

Jin W, Melo JR, Nagaki K, Talbert PB, Henikoff S, Dawe RK, Jiang J (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. Plant Cell 16:571–581

Karpen GH, Spradling AC (1990) Reduced DNA polytenization of a minichromosome region undergoing position-effect variegation in Drosophila. Cell 63:97–107

Karpen GH, Spradling AC (1992) Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. Genetics 132:737–753

Kato A, Lamb JC, Birchler JA (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci USA 101:13554–13559

Kikuchi Y (1983) Yeast plasmid requires a cis-acting locus and two plasmid proteins for its stable maintenance. Cell 35:487–493

Kuhn RM, Clarke L, Carbon J (1991) Clustered tRNA genes in *Schizosaccharomyces pombe* centromeric DNA sequence repeats. Proc Natl Acad Sci USA 88:1306–1310

Larin Z, Fricker MD, Tyler-Smith C (1994) De novo formation of several features of a centromere following introduction of a Y alphoid YAC into mammalian cells. Hum Mol Genet 3:689–695

Le MH, Duricka D, Karpen GH (1995) Islands of complex DNA are widespread in Drosophila centric heterochromatin. Genetics 141:283–303

Lechner J, Carbon J (1991) A 240 kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere. Cell 64:717–725

Lo AW, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E, Magliano DJ, Choo KH (2001a) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. EMBO J 20:2087–2096

Lo AW, Magliano DJ, Sibson MC, Kalitsis P, Craig JM, Choo KH (2001b) A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. Genome Res 11:448–457

Maddox PS, Hyndman F, Monen J, Oegema K, Desai A (2007) Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. J Cell Biol 176:757–763

Maggert KA, Karpen GH (2001) The activation of a neocentromere in Drosophila requires proximity to an endogenous centromere. Genetics 158:1615–1628

Maine GT, Surosky RT, Tye BK (1984) Isolation and characterization of the centromere from chromosome V (CEN5) of *Saccharomyces cerevisiae*. Mol Cell Biol 4:86–91

Mann C, Davis RW (1986) Structure and sequence of the centromeric DNA of chromosome 4 in *Saccharomyces cerevisiae*. Mol Cell Biol 6:241–245

Marshall OJ, Chueh AC, Wong LH, Choo KH (2008) Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am J Hum Genet 82:261–282

Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J Cell Biol 109:1963–1973

Masumoto H, Ikeno M, Nakano M, Okazaki T, Grimes B, Cooke H, Suzuki N (1998) Assay of centromere function using a human artificial chromosome. Chromosoma 107:406–416

McClintock B (1939) The behavior in successive nuclear divisions of a chromosome broken at meiosis. Proc Natl Acad Sci USA 25:405–416

McClintock B (1941) The stability of broken ends of chromosomes in *Zea mays*. Genetics 26:234–282

McGrew J, Diehl B, Fitzgerald-Hayes M (1986) Single base-pair mutations in centromere element III cause aberrant chromosome segregation in *Saccharomyces cerevisiae*. Mol Cell Biol 6:530–538

Measday V, Hailey DW, Pot I, Givan SA, Hyland KM, Cagney G, Fields S, Davis TN, Hieter P (2002) Ctf3p, the Mis6 budding yeast homolog, interacts with Mcm22p and Mcm16p at the yeast outer kinetochore. Genes Dev 16:101–113

Meluh PB, Yang P, Glowczewski L, Koshland D, Smith MM (1998) Cse4p is a component of the core centromere of *Saccharomyces cerevisiae*. Cell 94:607–613

Mizuguchi G, Xiao H, Wisniewski J, Smith MM, Wu C (2007) Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. Cell 129:1153–1164

Monen J, Maddox PS, Hyndman F, Oegema K, Desai A (2005) Differential role of CENP-A in the segregation of holocentric C. elegans chromosomes during meiosis and mitosis. Nat Cell Biol 7:1248–1255

Moreno-Moreno O, Torras-Llort M, Azorin F (2006) Proteolysis restricts localization of CID, the centromere-specific histone H3 variant of Drosophila, to centromeres. Nucleic Acids Res 34:6247–6255

Motamedi MR, Verdel A, Colmenares SU, Gerber SA, Gygi SP, Moazed D (2004) Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell 119:789–802

Mroczek RJ, Dawe RK (2003) Distribution of retroelements in centromeres and neocentromeres of maize. Genetics 165:809–819

Mroczek RJ, Melo JR, Luce AC, Hiatt EN, Dawe RK (2006) The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. Genetics 174:145–154

Murakami S, Matsumoto T, Niwa O, Yanagida M (1991) Structure of the fission yeast centromere cen3: direct analysis of the reiterated inverted region. Chromosoma 101:214–221

Murphy M, Fitzgerald-Hayes M (1990) Cis- and trans-acting factors involved in centromere function in *Saccharomyces cerevisiae*. Mol Microbiol 4:329–336

Murphy MR, Fowlkes DM, Fitzgerald-Hayes M (1991) Analysis of centromere function in *Saccharomyces cerevisiae* using synthetic centromere mutants. Chromosoma 101:189–197

Murphy TD, Karpen GH (1995) Localization of centromere function in a Drosophila minichromosome. Cell 82:599–609

Murzina NV, Pei XY, Zhang W, Sparkes M, Vicente-Garcia J, Pratap JV, McLaughlin SH, Ben-Shahar TR, Verreault A, Luisi BF, Laue ED (2008) Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. Structure 16:1077–1085

Nakano M, Cardinale S, Noskov VN, Gassmann R, Vagnarelli P, Kandels-Lewis S, Larionov V, Earnshaw WC, Masumoto H (2008) Inactivation of a human kinetochore by specific targeting of chromatin modifiers. Dev Cell 14:507–522

Nakaseko Y, Adachi Y, Funahashi S, Niwa O, Yanagida M (1986) Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. EMBO J 5:1011–1021

Nakayama J, Klar AJ, Grewal SI (2000) A chromodomain protein, Swi6, performs imprinting functions in fission yeast during mitosis and meiosis. Cell 101:307–317

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292:110–113

Nasuda S, Hudakova S, Schubert I, Houben A, Endo TR (2005) Stable barley chromosomes without centromeric repeats. Proc Natl Acad Sci USA 102:9842–9847

Neitz M, Carbon J (1985) Identification and characterization of the centromere from chromosome XIV in *Saccharomyces cerevisiae*. Mol Cell Biol 5:2887–2893

Neumann P, Yan H, Jiang J (2007) The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. Genetics 176:749–761

Noma K, Allis CD, Grewal SI (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science 293:1150–1155

Obuse C, Yang H, Nozaki N, Goto S, Okazaki T, Yoda K (2004) Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. Genes Cells 9:105–120

Ohzeki J, Nakano M, Okada T, Masumoto H (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J Cell Biol 159:765–775

Okada M, Cheeseman IM, Hori T, Okawa K, McLeod IX, Yates JR III, Desai A, Fukagawa T (2006) The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. Nat Cell Biol 8:446–457

Okada T, Ohzeki J, Nakano M, Yoda K, Brinkley WR, Larionov V, Masumoto H (2007) CENP-B controls centromere formation depending on the chromatin context. Cell 131:1287–1300

Ortiz J, Stemmann O, Rank S, Lechner J (1999) A putative protein complex consisting of Ctf19, Mcm21, and Okp1 represents a missing link in the budding yeast kinetochore. Genes Dev 13:1140–1155

Palmer DK, O'Day K, Margolis RL (1989) Biochemical analysis of CENP-A, a centromeric protein with histone-like properties. Prog Clin Biol Res 318:61–72

Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci USA 88:3734–3738

Palmer DK, O'Day K, Wener MH, Andrews BS, Margolis RL (1987) A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J Cell Biol 104:805–815

Panzeri L, Philippsen P (1982) Centromeric DNA from chromosome VI in *Saccharomyces cerevisiae* strains. EMBO J 1:1605–1611

Partridge JF, Borgstrom B, Allshire RC (2000) Distinct protein interaction domains and protein spreading in a complex centromere. Genes Dev 14:783–791

Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ (1981) Highly repeated DNA sequence limited to knob heterochromatin in maize. Proc Natl Acad Sci USA 78:4490–4494

Pidoux AL, Richardson W, Allshire RC (2003) Sim4: a novel fission yeast kinetochore protein required for centromeric silencing and chromosome segregation. J Cell Biol 161:295–307

Polizzi C, Clarke L (1991) The chromatin structure of centromeres from fission yeast: differentiation of the central core that correlates with function. J Cell Biol 112:191–201

Ray-Gallet D, Quivy JP, Scamps C, Martini EM, Lipinski M, Almouzni G (2002) HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. Mol Cell 9:1091–1100

Rhoades MM, Vilkomerson H (1942) On the anaphase movement of chromosomes. Proc Natl Acad Sci USA 28:433–436

Rivera H, Vassquez AI, Ayala-Madrigal ML, Ramirez-Duenas ML, Davalos IP (1996) Alphoidless centromere of a familial unstable inverted Y chromosome. Ann Genet 39:236–239

Saffery R, Irvine DV, Griffiths B, Kalitsis P, Wordeman L, Choo KH (2000) Human centromeres and neocentromeres show identical distribution patterns of >20 functionally important kinetochore-associated proteins. Hum Mol Genet 9:175–185

Saffery R, Wong LH, Irvine DV, Bateman MA, Griffiths B, Cutts SM, Cancilla MR, Cendron AC, Stafford AJ, Choo KH (2001) Construction of neocentromere-based human minichromosomes by telomere-associated chromosomal truncation. Proc Natl Acad Sci USA 98:5705–5710

Saitoh S, Takahashi K, Yanagida M (1997) Mis6, a fission yeast inner centromere protein, acts during G1/S and forms specialized chromatin required for equal segregation. Cell 90:131–143

Schueler MG, Sullivan BA (2006) Structural and functional dynamics of human centromeric chromatin. Annu Rev Genomics Hum Genet 7:301–313

Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. Science 294:109–115

Schuh M, Lehner CF, Heidmann S (2007) Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. Curr Biol 17:237–243

Scott-Drew S, Murray JA (1998) Localisation and interaction of the protein components of the yeast 2 mu circle plasmid partitioning system suggest a mechanism for plasmid inheritance. J Cell Sci 111(Pt 13):1779–1789

Shelby RD, Monier K, Sullivan KF (2000) Chromatin assembly at kinetochores is uncoupled from DNA replication. J Cell Biol 151:1113–1118

Smith S, Stillman B (1989) Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. Cell 58:15–25

Som T, Armstrong KA, Volkert FC, Broach JR (1988) Autoregulation of 2 micron circle gene expression provides a model for maintenance of stable plasmid copy levels. Cell 52:27–37

Sorger PK, Severin FF, Hyman AA (1994) Factors required for the binding of reassembled yeast kinetochores to microtubules in vitro. J Cell Biol 127:995–1008

Steiner NC, Clarke L (1994) A novel epigenetic effect can alter centromere function in fission yeast. Cell 79:865–874

Steiner NC, Hahnenberger KM, Clarke L (1993) Centromeres of the fission yeast *Schizosaccharomyces pombe* are highly variable genetic loci. Mol Cell Biol 13:4578–4587

Stinchcomb DT, Struhl K, Davis RW (1979) Isolation and characterisation of a yeast chromosomal replicator. Nature 282:39–43

Stinchcomb DT, Mann C, Davis RW (1982) Centromeric DNA from *Saccharomyces cerevisiae*. J Mol Biol 158:157–190

Stoler S, Keith KC, Curnick KE, Fitzgerald-Hayes M (1995) A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev 9:573–586

Stoler S, Rogers K, Weitze S, Morey L, Fitzgerald-Hayes M, Baker RE (2007) Scm3, an essential *Saccharomyces cerevisiae* centromere protein required for G2/M progression and Cse4 localization. Proc Natl Acad Sci USA 104:10571–10576

Sugata N, Li S, Earnshaw WC, Yen TJ, Yoda K, Masumoto H, Munekata E, Warburton PE, Todokoro K (2000) Human CENP-H multimers colocalize with CENP-A and CENP-C at active centromere--kinetochore complexes. Hum Mol Genet 9:2919–2926

Sullivan BA, Karpen GH (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat Struct Mol Biol 11:1076–1083

Sullivan BA, Schwartz S (1995) Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. Hum Mol Genet 4:2189–2197

Sullivan BA, Willard HF (1998) Stable dicentric X chromosomes with two functional centromeres. Nat Genet 20:227–228

Sullivan KF, Hechenberger M, Masri K (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol 127:581–592

Sun X, Wahlstrom J, Karpen G (1997) Molecular structure of a functional Drosophila centromere. Cell 91:1007–1019

Sun X, Le HD, Wahlstrom JM, Karpen GH (2003) Sequence analysis of a functional Drosophila centromere. Genome Res 13:182–194

Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y (2004) Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell 116:51–61

Takahashi K, Murakami S, Chikashige Y, Funabiki H, Niwa O, Yanagida M (1992) A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. Mol Biol Cell 3:819–835

Takahashi K, Chen ES, Yanagida M (2000) Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. Science 288:2215–2219

Takayama Y, Sato H, Saitoh S, Ogiyama Y, Masuda F, Takahashi K (2008) Biphasic incorporation of centromeric histone CENP-A in fission yeast. Mol Biol Cell 19:682–690

Topp CN, Zhong CX, Dawe RK (2004) Centromere-encoded RNAs are integral components of the maize kinetochore. Proc Natl Acad Sci USA 101:15986–15991

Tower J, Karpen GH, Craig N, Spradling AC (1993) Preferential transposition of Drosophila P elements to nearby chromosomal sites. Genetics 133:347–359

Tyler-Smith C, Gimelli G, Giglio S, Floridia G, Pandya A, Terzoli G, Warburton PE, Earnshaw WC, Zuffardi O (1999) Transmission of a fully functional human neocentromere through three generations. Am J Hum Genet 64:1440–1444

Valdivia MM, Brinkley BR (1985) Fractionation and initial characterization of the kinetochore from mammalian metaphase chromosomes. J Cell Biol 101:1124–1134

Ventura M, Weigl S, Carbone L, Cardone MF, Misceo D, Teti M, D'Addabbo P, Wandall A, Bjorck E, de Jong PJ, She X, Eichler EE, Archidiacono N, Rocchi M (2004) Recurrent sites for new centromere seeding. Genome Res 14:1696–1703

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303:672–676

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297:1833–1837

Voullaire LE, Slater HR, Petrovic V, Choo KH (1993) A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere. Am J Hum Genet 52:1153–1163

Warburton PE, Cooke HJ (1997) Hamster chromosomes containing amplified human alpha-satellite DNA show delayed sister chromatid separation in the absence of de novo kinetochore formation. Chromosoma 106:149–159

Warburton PE, Cooke CA, Bourassa S, Vafa O, Sullivan BA, Stetten G, Gimelli G, Warburton D, Tyler-Smith C, Sullivan KF, Poirier GG, Earnshaw WC (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr Biol 7:901–904

Warburton PE, Dolled M, Mahmood R, Alonso A, Li S, Naritomi K, Tohma T, Nagai T, Hasegawa T, Ohashi H, Govaerts LC, Eussen BH, Van Hemel JO, Lozzio C, Schwartz S, Dowhanick-Morrissette JJ, Spinner NB, Rivera H, Crolla JA, Yu C, Warburton D (2000) Molecular cytogenetic analysis of eight inversion duplications of human chromosome 13q that each contain a neocentromere. Am J Hum Genet 66:1794–1806

Westermann S, Cheeseman IM, Anderson S, Yates JR III, Drubin DG, Barnes G (2003) Architecture of the budding yeast kinetochore reveals a conserved molecular core. J Cell Biol 163:215–222

Willard HF (1991) Evolution of alpha satellite. Curr Opin Genet Dev 1:509–514

Williams BC, Murphy TD, Goldberg ML, Karpen GH (1998) Neocentromere activity of structurally acentric mini-chromosomes in Drosophila. Nat Genet 18:30–37

Worcel A, Han S, Wong ML (1978) Assembly of newly replicated chromatin. Cell 15:969–977

Wu RS, Tsai S, Bonner WM (1982) Patterns of histone variant synthesis can distinguish G0 from G1 cells. Cell 31:367–374

Yu HG, Hiatt EN, Chan A, Sweeney M, Dawe RK (1997) Neocentromere-mediated chromosome movement in maize. J Cell Biol 139:831–840

Zhang W, Lee HR, Koo DH, Jiang J (2008) Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in Arabidopsis thaliana and maize. Plant Cell 20:25–34

Zheng YZ, Roseman RR, Carlson WR (1999) Time course study of the chromosome-type breakage-fusion-bridge cycle in maize. Genetics 153:1435–1444

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

Zinkowski RP, Meyne J, Brinkley BR (1991) The centromere-kinetochore complex: a repeat subunit model. J Cell Biol 113:1091–1110

Chapter 2
The Centromere-Drive Hypothesis: A Simple Basis for Centromere Complexity

Harmit S. Malik

Contents

Abstract Centromeres are far more complex and evolutionarily labile than expected based on their conserved, essential function. The rapid evolution of both centromeric DNA and proteins strongly argue that centromeres are locked in an evolutionary conflict to increase their odds of transmission during asymmetric (female) meiosis. Evolutionary success for "cheating" centromeres can result in highly deleterious consequences for the species, either in terms of skewed sex ratios or male sterility. Centromeric proteins evolve rapidly to suppress the deleterious effects of "centromere-drive." This chapter summarizes the mounting evidence in favor of the centromere-drive model, and its implications for centromere evolution in taxa with variations in meiosis.

H.S. Malik
Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, Seattle, WA, 98109
e-mail: hsmalik@fhcrc.org

Đ. Ugarković (ed.), *Centromere,* Progress in Molecular and Subcellular Biology 48,
DOI: 10.1007/978-3-642-00182-6_2, © Springer-Verlag Berlin Heidelbarg 2009

2.1 Centromere Complexity in Eukaryotes

Centromeres provide a universal means to faithfully segregate chromosomes in eukaryotes. They are the chromosomal sites that act as binding sites for microtubules that mediate the mechanical force that pulls chromosomes or chromatids apart during meiosis and mitosis. Despite this conserved function, centromeres can dramatically range in size and complexity. The simplest centromeres are the 125 bp point centromeres in *Saccharomyces cerevisiae* (Fitzgerald-Hayes et al. 1982). More complex centromeres are found in fission yeast *Schizosaccharomyces pombe* (Clarke and Baum 1990; Wood et al. 2002). In contrast, centromeres in plants and animals are highly complex and consist of hundreds of kilobases of long arrays of satellite repeats (Copenhaver et al. 1999; Schueler et al. 2001). A further degree of complexity is evident in the centromeres of holokinetic organisms like *Caenorhabditis elegans;* centromeric determinants dispersed throughout the length of the chromosome that coalesce at metaphase, such that each centromere runs the entire length of the chromosome (Buchwitz et al. 1999) although it appears that meiotic chromosome segregation may be dramatically different from mitosis in such cases (Monen et al. 2005). On the other end of the spectrum from holokinetic organisms are human neocentromeres, which appear to lack any tandemly repetitive sequence whatsoever (Lo et al. 2001). In *Drosophila melanogaster*, centromeric satellites can be found in distal blocks from the centromeres, some of which have weak centromeric activity (Platero et al. 1999). Centromeric and heterochromatic sequences are almost indistinguishable in the best studied Drosophila centromere (Sun et al. 2003). Similarly, in the human genome, it is unclear what subset of α-satellites are centromeric versus heterochromatic.

There are large technical challenges associated with sequencing and assembling highly repetitive centromeric regions in eukaryotes. Despite this, a picture of centromere complexity and the events that shape their evolution has emerged from herculean sequencing and assembly efforts in diverse organisms. The 420 kb long Dpl187 minichromosome in *D. melanogaster* (Sun et al. 2003), the 750 kb centromere on rice chromosome 8 (Nagaki et al. 2004), and the human X centromere (Schueler et al. 2001) are examples of assembly efforts that have led to a detailed picture of the heterochromatin-centromere boundary in complex centromeres. For instance, the assembly of the human X centromere indicated a highly homogeneous region of α-satellite repeats at the "core" of centromeres. This core is flanked by satellite repeats with a gradient of increased heterogeneity (accumulated mutations) and transposon insertions with physical distance away from the core. Analysis of mutations and insertions in the flanking region led to the surprising model that the extant X centromere α-satellite is young and probably arose only in the great apes (Schueler et al. 2001). These studies are still akin to looking at the "flotsam on the beach" (the boundaries of centromeres) and to decipher what the "middle of the ocean" (the homogeneous centromeric satellites in the middle of the array) might look like.

Nevertheless, these findings have been instructive. They support the simple mutation-recombination balance model where recombination (either unequal crossing over or gene conversion) is the underlying force that homogenizes centromeric

repeats in the middle of an array, balanced by mutation and transposition in the flanks (Malik and Henikoff 2002; McAllister and Werren 1999, see Chap. 3 in this volume). Adding to the complexity of centromeric regions *within* a species is the finding that satellite DNA sequences can change quite rapidly *between* closely related species. For instance, there is very little overlap between the centromeric satellite sequences of *Drosophila melanogaster* and *D. simulans*, in spite of the fact that many satellites are shared between the two species (Lohe and Brutlag 1987). Satellite repeats in rice centromeres also have been found to have dramatically changed over short evolutionary periods (Lee et al. 2005; see Chap. 6 in this volume). Similarly, the human X centromeric satellite appears to be only as old as the great apes (Schueler et al. 2001). In several instances, homologous chromosomes in closely related primate species bear different, non-orthologous α-satellite sequence variants (Haaf and Willard 1997; Samonte et al. 1997). Thus, centromeric regions evolve rapidly between species.

2.2 Rapid Evolution of Centromeres is not Due to Relaxed Selective Constraint

Studies on centromeric DNA paint a highly dynamic picture of centromere evolution, but they do not provide a rationale for this rapid evolution and large-scale accumulation of satellite repeats. Indeed, several theoretical studies have pointed out the inadequacy of mutation and recombination alone to explain increased array sizes, suggesting that selection must play a role in their evolution (Charlesworth et al. 1994; Stephan 1989; Stephan and Cho 1994; Walsh 1987). There is precedent for the view that alterations in numbers and sequences of DNA-satellite repeats can have fitness consequences. Pericentric satellites have been shown to contribute to a fitness difference within *D. melanogaster* strains (Wu et al. 1989). Yet another pericentric satellite contributes to hybrid inviability between *D. simulans/D. melanogaster* interspecific hybrids (Sawamura and Yamamoto 1993; Sawamura et al. 1993).

In human centromeres, α-satellites are organized at the centromeres into two types of repeat structures. At the "central core" of centromeric regions, α-satellites are found in a repeat unit that consists of multiple monomers. This multi-monomer unit is repeated over and over to make up a higher-order array. Higher-order arrays of α-satellite are the typical sequence organization of centromere regions of humans and can stretch for megabases of DNA that is largely uninterrupted by any kind of insertion or mutation. For example, the repeat unit length of the central core of the human X chromosome is ~2 kb, comprised of twelve 171-bp monomers of an evolutionarily young DXZ1 α-satellite (Schueler et al. 2001). Surrounding this "central region" are α-satellites found in monomeric units. These α-satellites are considered pericentric; while they may serve important roles in chromosome segregation, they do not recruit centromeric and kinetochore proteins. It is highly likely that monomeric α-satellite structures actually represent the ancestral state of the primate centromeres. Present-day heterochromatic α-satellites might be an evolutionary relic of ancestrally centromeric α-satellite that have lost centromeric function, and accreted to the edges

of the array, acquiring mutations and transposon insertion events, whereas the "central core" is cleansed of mutations and insertions by recombination.

What are the selective constraints that might act on centromeric DNA? One form of selection could be simply purifying selection to maintain an uninterrupted, homogeneous array of a minimum size, so that it can form a functional centromere (e.g., the higher order uninterrupted array of DXZ1 satellites on the human X chromosome). This can explain the highly homogeneous centromeric satellites found at the core of most centromeres. One can evaluate additional selective constraints acting on centromeric DNA by comparisons of the centromeric central core to the pericentric monomer units. The pericentric monomers provide a good yardstick for this comparison because they are presumed to be selectively neutral (or nearly so). Comparison of the monomeric units and centromeric higher-order array units from orthologous chromosomes (e.g., chimp vs. human) leads to the surprising finding that the centromeric arrays from different species are more divergent than the pericentric units (Rudd et al. 2006). These findings are counter-intuitive because the centromeric α-satellite array is the functional centromere and is under stringent selective constraint, while the pericentric α-satellites are not.

It is important to point out that there is no *a priori* expectation that satellite repeats should evolve faster than nonrepetitive DNA in the absence of any biases introduced by selection. This is because mutations in any particular satellite repeat (introduced with a mutation rate, μ) have a probability of fixation that is proportional to their initial incidence ($1/2N$, where $2N$ are the number of repeat units in arrays on both homologous chromosomes). Thus, the overall likelihood for any mutation spreading to fixationsummed over the entire array equals $2N$ times $\mu/2N$, which equals the mutation rate (μ) for nonrepetitive DNA.

In sum, one is left with the paradoxical observation that the satellite units that are most constrained *within* a species have evolved most rapidly *between* species. It is this paradoxical observation that leads to the idea that some selective force must actively drive the rapid fixation of mutations at centromeric satellites by imposing a bias in favor of retaining mutations. Intriguingly, rapid evolution of centromeric satellites is not only seen in primate and Drosophila centromeres, but also dramatically in the case of plant centromeres (Lee et al. 2005). A recent study has also found that budding yeast centromeric DNA is one of the most rapidly evolving components of the *S. cerevisiae* genome, although here it is more likely that increased mutational rate at the centromere is the likely explanation (Bensasson et al. 2008).

2.3 Centromeric Histones Epigenetically Define Centromeres in Most Eukaryotes

The notion that centromeres are epigenetically, and not genetically, defined in most eukaryotes is the subject of Chap. 1 of this volume. Instead of covering all the evidence in favor of the epigenetic model here, the reader is directed to that chapter

for all the pertinent information. In this chapter, we focus on the likely "mark" of epigenetically defining centromeres: the centromeric histone variant and nucleosomes bearing this variant. Centromeric histones (CenH3s) are variant members of the Histone H3 family of proteins. Initially discovered as the CENP-A protein in mammals (Palmer et al. 1987), CenH3s are now found to be encoded by a single gene in every eukaryotic genome studied so far (Malik and Henikoff 2003) and are essential for accurate chromosome segregation (Blower and Karpen 2001; Buchwitz et al. 1999; Stoler et al. 1995). They substitute for canonical H3 in variant nucleosomes (Sullivan et al. 1994; Yoda et al. 2000, 2004) and their localization can discriminate between the centromere and the surrounding heterochromatin (Takahashi et al. 2000). Thus, CenH3s provide a faithful marker of centromere identity throughout the entire range of centromere sizes, from the point centromeres of *S. cerevisiae* to the holokinetic centromeres of *C. elegans* (Buchwitz et al. 1999; Stoler et al. 1995).

2.3.1 Distinguishing Features of Centromeric Histones

Centromeric histones differ from canonical histones in four key sequence features, highlighted in Fig. 2.1 (Malik and Henikoff 2003; Shelby et al. 1997; Sullivan et al. 1994). First, while canonical H3s in all eukaryotes have a well-conserved N-terminal tail, the N-terminal tails of CenH3s vary in both length and sequence and cannot be aligned across different lineages (Fig. 2.1a). Second, all CenH3s have a longer Loop1 region than canonical H3s. Loop1 is one of the principal DNA-interaction domains for H3 (Luger et al. 1997), and the longer Loop1 of CenH3s has been inferred to allow them a greater DNA-binding specificity (Shelby et al. 1997). Recent studies have also firmly established that Loop1 and helix α2 of the CenH3s together represent a centromeric targeting domain (CATD), which specifically distinguishes CenH3s from canonical histone H3 (Black et al. 2004). Indeed, making a chimeric H3 that possesses a CATD from a CenH3 is capable of localizing and functioning in mitosis in both budding yeast and human cells (see Chap. 1) (Black et al. 2007). Finally, in a comparison of just the core histone fold domains (HFD), we found that CenH3s appear to have evolved more rapidly in contrast to canonical histone H3 (Henikoff et al. 2001; Malik and Henikoff 2003) (Fig. 2.1b). This suggests either that CenH3s are less constrained than canonical H3 or that they are subject to rapid evolution (see Sect. 4.2.2).

Recent studies indicate that CenH3-containing nucleosomes are present in distinct blocks interspersed with blocks of canonical H3-containing nucleosomes (Ahmad and Henikoff 2002; Blower et al. 2002) (Fig. 2.1c). The proportion of centromeric DNA packaged by CenH3-nucleosomes is likely determined by the dynamics and affinity of CenH3 vs. canonical H3 nucleosomes for binding (Blower et al. 2002; Nagaki et al. 2004). For instance, over-expression of heterochromatin proteins can encroach onto centromeric DNA and affect chromosome segregation (Halverson et al. 1997, 2000). This suggests that CenH3s epigenetically delineate

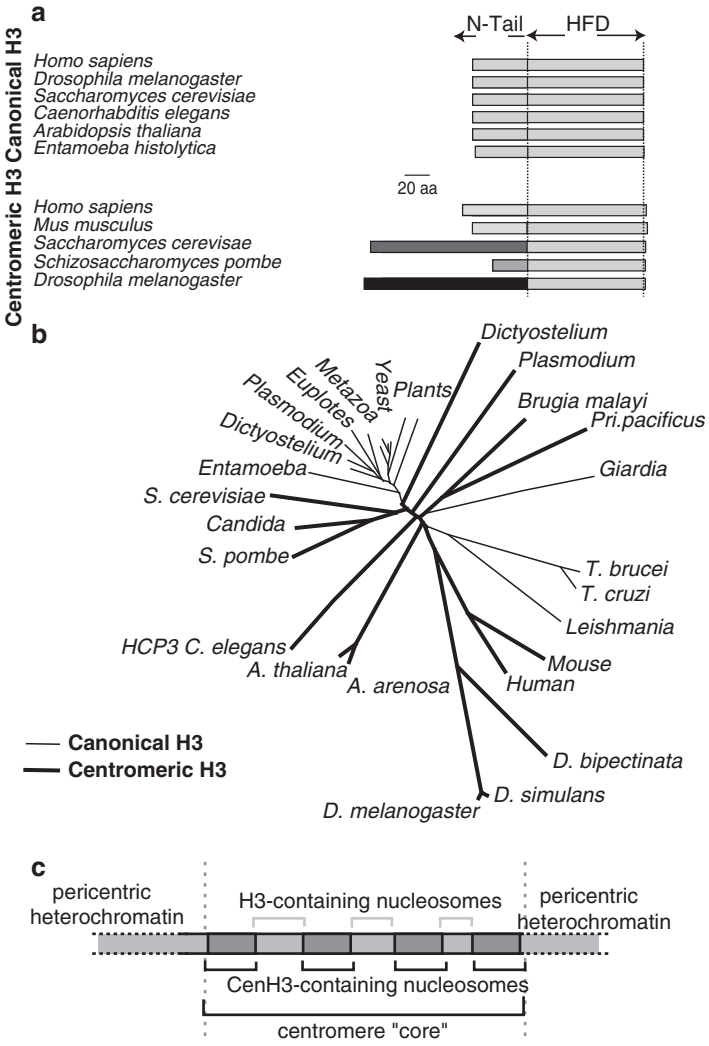


Fig. 2.1 Comparison of canonical and centromeric H3 proteins (Henikoff et al. 2001). (**a**) Canonical and centromeric H3 proteins showing that the N-terminal tail in CenH3s is not as well conserved as in canonical H3s. (**b**) Neighbor-joining phylogeny of the HFD domains indicates that CenH3s are more rapidly evolving (longer branch lengths). (**c**) CenH3 and canonical nucleosomes are present in interspersed blocks in the centromere "core," but pericentric heterochromatin and euchromatin (not shown) are packaged exclusively by canonical nucleosomes (Blower et al. 2002)

centromere boundaries and may define centromere strength. Furthermore, their association with centromeres is highly dynamic and dependent on the relative DNA-binding affinities of CenH3s, canonical histones, as well as satellite-binding proteins. Modulating the DNA-binding affinity of any one of these entities may affect centromere size and strength.

2.3.2 Centromeric Histones Evolve Rapidly in Drosophila

To dissect the selective constraints acting on centromeric histones in a more fine-scale analysis than in Fig. 2.1, we focused on the Drosophila CenH3 gene, *Cid* (for *centromere identifier*). We compared the Cid coding sequences from multiple geographical strains of *D. melanogaster* and *D. simulans* to an outgroup, *D. teissieri*, and parsed all the changes into two separate categories. The first category separated changes that caused a change in the amino acid being encoded by the codon (Replacement) from those that did not (Synonymous). The second category separated changes that were fixed in either species, after separation from a common ancestor, from those that were polymorphic within either species. Under the model for neutral evolution, Rf:Sf should approximate Rp:Sp, whereas finding an excess of Rf changes would suggest that many of these replacement changes were fixed due to an adaptive advantage (positive selection) (McDonald and Kreitman 1991). In our *Cid* analysis, we found that Rf:Sf and Rp:Sp were 18:10 and 9:28, respectively. Under the neutral model, we would have expected only ~3 Rf changes ($9/28 \times 10$) but found 18 instead (Malik and Henikoff 2001). These findings reject the neutral evolution model with high confidence ($p < 0.0025$) and support the finding that *Cid* has been subject to positive selectionin Drosophila.

Furthermore, we could show that both *D. melanogaster* and *D. simulans* Cid were subject to positive selection in their N-terminal tail and histone fold domain (HFD), respectively (see Fig. 2.2). In the case of *D. melanogaster*, we found evidence for a recent adaptive sweep that reduced the synonymous polymorphisms in the N-terminal tail by a Hudson–Kreitman–Aguade test ($p < 0.05$) (Hudson et al. 1987).

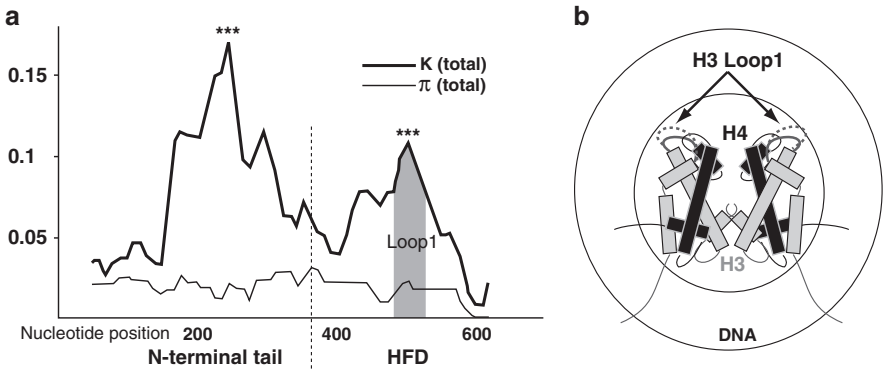


Fig. 2.2 Cid polymorphisms (Malik and Henikoff 2001). (**a**) A sliding window analysis of the intraspecific polymorphism in *D. simulans* represented by π and the interspecific divergence (K) for *Cid* performed using all sites (synonymous and replacement), with the *x*-axis indicating nucleotide position. The *dashed line* separates the N-terminal tail region from the C-terminal HFD, with the *shaded area* indicating the Loop1 region. Both the N-terminal tail and the HFD have an excess of fixed replacements (***) in a McDonald–Kreitman test. (**b**) Nucleosomal structure with H3, H4, and DNA (H2A and H2B are not shown for clarity) highlights the Loop1 region of H3 (Luger et al. 1997). CenH3 Loop1 is longer but occupies a similar position

The positive selection in the HFD domain could be mapped onto the crystal structure of the nucleosome, as Cid was ~65% identical to H3 in amino acid alignments. We found that all the fixed replacement changes occurred in a very small segment of the HFD that corresponded to the Loop1 region of Cid (Fig. 2.2), suggesting that altered DNA-binding specificity was driving the positive selection of this essential gene for chromosome segregation (see Sect. 4.2.3).

2.3.3 *Rapid Evolutionary Changes in Loop1 have Dramatic Functional Consequences*

Our evolutionary analyses identified recurrent episodes of positive selection in the Loop1 region of Cid. In parallel experiments, we assayed whether the rapid evolution of *Cid* relative to canonical H3 translated to any gross effect in terms of centromere function or targeting (Vermaak et al. 2002). We assayed for centromere targeting of divergent *Cid* genes by introducing GFP-tagged versions of *Cid* from a variety of Drosophila species by transient transfection in Kc tissue culture cells (Fig. 2.3).

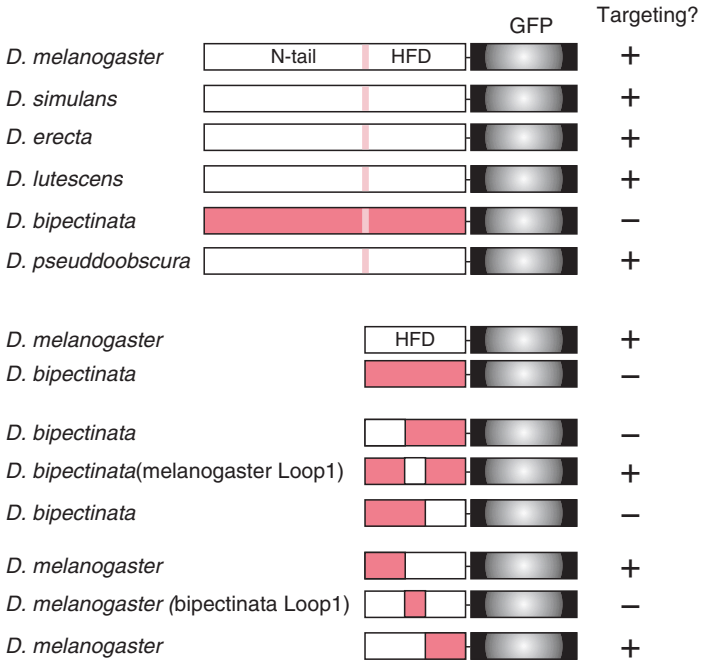


Fig. 2.3 Localization of Cid from divergent *Drosophila* species in *D. melanogaster* tissue culture cells (Vermaak et al. 2002). GFP-tagged *Cid* genes from Drosophila species, representing increased evolutionary distances, were introduced into Kc cells. Of these, only *D. bipectinata Cid* did not localize correctly to centromeres. The HFD domain was necessary and sufficient for the targeting (not shown), and this targeting was completely dependent on the Loop1 region of the HFD

The endogenous *D. melanogaster Cid* was assayed using a specific antibody, while the introduced genes were assayed by GFP localization. *Cid* genes from *D. melanogaster, D. simulans, D. erecta, D. lutescens*, and *D. pseudoobscura* targeted appropriately to *D. melanogaster* centromeres, whereas Cid from *D. bipectinata* (*ananassae* subgroup) did not. This centromere targeting ability was dependent on HFD alone. *D. melanogaster* Cid-HFD targeted appropriately to centromeres in Kc cells, whereas Cid-HFD from *D. bipectinata* did not.

In chimeric swaps between segments of the *D. melanogaster* and *D. bipectinata* HFD domains, replacing the *D. bipectinata* Loop1 region from *D. melanogaster* restored centromere targeting to the chimera. Even more strikingly, replacing the *D. melanogaster* Loop1 region with that from *D. bipectinata* abrogated targeting. These targeting experiments showed that the Loop1 region is critical for targeting Cid appropriately to centromeric DNA. Further site-directed mutation analysis of Loop1 also revealed that several residues in the Loop1 region that found under purifying selection previously (Fig. 2.3a) were also important for mediating correct targeting. This suggests that Loop1 contains both conserved and positively selected residues that are required for correctly targeting Cid to centromeres. Notably, the differences of centromere targeting between Cid genes from two species of Drosophila less than 25 million years diverged from each other, also argue that the CATD domain (a distinguishing feature of CenH3s separate from canonical H3s) has even functionally altered within a lineage of CenH3s (Vermaak et al. 2002). Thus, changes that we identified in CenH3s as being driven by positive selection were functionally important for the correct localization and functioning of CenH3s.

2.3.4 Centromeric Protein Evolution Outside Drosophila

Similar results for the rapid evolution of CenH3s were also seen in the case of the Arabidopsis CenH3, HTR12 (Talbert et al. 2002), and even here there was a strong focus of positive selection acting on the Loop1 region (Cooper and Henikoff 2004). Thus, in both plants and animals, it appears that the single, essential centromeric histone gene that defines the epigenetic basis of centromeres has been subject to the types of selective pressures typically seen only in cases of rapid adaptation. Intriguingly, the initial findings of positive selection acting on centromeric histones have also been extended to a second, ubiquitously found, essential centromeric protein, CENP-C (Talbert et al. 2004). In fact, it turns out that CENP-C provides more consistent signatures of rapid evolution; in mammalian genomes, for instance, CENP-C but not CENP-A (vertebrate CenH3) shows evidence of positive selection. It is unclear what differences in selective constraint drive the positive selection of CenH3s in flies and plants to be different from mammals. One possibility is that CENP-A localization and DNA-binding preferences are dictated by another protein, perhaps a chaperone, whereas this is not the case for Cid or HTR12.

However, in budding yeasts like *S. cerevisiae*, no evidence of positive selection was seen in the CENP-C gene, *Mif2*, or the CenH3 gene, *Cse4*. Intriguingly, this

lineage of yeasts is also atypical in eukaryotes for having small, genetically defined "point" centromeres. This suggested that a fundamentally altered process of chromosome segregation may be influencing the rapid evolution of centromeric components in animals and plants, but not budding yeasts. These dual signatures of rapid evolution in centromeric DNA (Sect. 4.1) and centromeric proteins (Sect. 4.2) are indicative of a genetic conflict constantly reshaping these components in plants and animals exclusively. We believe that asymmetric (female) meiosis is one distinguishing feature that is a common explanation for all these observations.

2.4 Asymmetry in Female Meiosis as a Driving Force in Evolutionary Biology

The asymmetric nature of female meiosis in plants and animals can lead to genetic elements subverting this process for their own advantage. The knob elements from maize are an example of such an entity (Rhoades 1942). Knobs are blocks of heterochromatin that are always found distally from the centromere. If a pair of chromosomes is heterozygous, that is, only one contains a knob, crossing over can occur between the knob and centromere during female meiosis. Under the appropriate genetic background, knobs bind microtubules and knob-bearing chromatids are pulled toward the outermost megaspores during Meiosis II. One of these outermost megaspores will become the gametophyte and produce gametes (Dawe and Cande 1996). Thus, instead of a 50% expected ratio of transmission in a heterozygote, knob transmission in female meiosis varies from 59 to 82% correlated with the size of the satellite array (Buckler et al. 1999). Thus, the "selfish" knobs exploit the inherently non-Mendelian nature of female meiosis for their survival.

A transmission advantage in female meiosis may also account for high rates of nondisjunction in Drosophila females (Zwick et al. 1999). A sensitized assay found a large range of nondisjunction frequencies among X chromosomes. This variation in nondisjunction correlated significantly with the two variants of the nod chromokinesin, which were found to be present at intermediate frequencies in natural populations. The nod chromokinesin is required for achiasmate segregation (Hawley et al. 1992; Karpen et al. 1996; Zhang et al. 1990), yet apparently deleterious alleles had thrived in Drosophila populations. These findings led to the oötid-competition model, which proposed that polymorphic alleles of loci involved in segregation of oötids during female meiosis were likely to provide multiple opportunities for competitive interactions among oötids, since only one oötid is included in the pronucleus (Zwick et al. 1999). Thus, female meiotic drive could result in the sponsoring of otherwise defective alleles, as a balance is struck between the competitive advantage conferred by this allele in female meiosis with its cost in causing high rates of nondisjunction. This model also predicted that centromeres and other chromosomal elements could compete directly in this manner, except that centromeres would competitively orient towards the preferred pole during Meiosis I, whereas telomeres and other distal elements would do so later in female meiosis (like the

knob elements in maize). This model serves as the basis of the "centromere-drive" model that we have proposed to explain the evolution of centromeres and their histones (Henikoff et al. 2001; Malik and Henikoff 2001).

Success in female meiosis may also negatively influence male meiosis. For example, Robertsonian fusions that result from the fusion of two acrocentric chromosomes have a differential advantage through female but not male meiosis in mice, humans, and chickens. In both humans and chickens, the Robertsonian fusions are preferentially transmitted through female meiosis, but in mice, it is the acrocentrics that are preferred (Pardo-Manuel de Villena and Sapienza 2001a, b). Thus, asymmetric female meiosis has great explanatory value in the evolution of mammalian karyotypes (mice have predominantly acrocentric chromosomes, whereas humans and birds have primarily metacentric karyotypes). A significant proportion (0.12%) of the human population are carriers of a Robertsonian translocation (Nielsen and Wohlert 1991). There are no reports of any somatic (mitotic) effects, but a significant fraction of male carriers of Robertsonian fusions appear to be partially-to-completely sterile (Daniel 2002). This sterility likely results from a male meiotic checkpoint that monitors tension of microtubule attachment in mice (Eaker et al. 2001) and may occur in Drosophila as well (McKee et al. 1998). Thus, female meiotic success can be balanced by the high cost to male fertility.

Under such a situation, where meiotic drivers have thrived in a population but cannot drive to fixation, theory predicts that suppressor alleles may arise to alleviate the effects of the drive or to eliminate the drive itself (Sandler and Novitski 1957). These suppressor alleles would be unlinked from the drive locus so as to not reap the "benefits" of the drive (Hartl 1975). Success of the suppressor alleles can lead to the degeneration of the drive system (in the absence of a transmission advantage) and subsequently to the degeneration of the suppressor, leading to the presence of *cryptic* drive-suppressor systems (Tao et al. 2001). Typically meiotic drivers and their suppressors are neomorphs (Merrill et al. 1999) and neither is essential for an organism. In the unusual scenario when essential elements act as drivers or suppressors, we could only uncover this *cryptic* genetic conflict by observing episodes of positive selection in them (Henikoff and Malik 2002).

2.5 Female Meiotic Drive vs. Male Post-Meiotic Dysfunction

The original proposal of meiotic drive (Sandler and Novitski 1957) was essentially a description of how asymmetric success in female meiotic drive could translate to differential evolutionary success. However, when we invoke the term "meiotic drive," typically most of the cases described are in fact concentrated on post-meiotic mechanisms. A celebrated example is the Segregation Distorter (SD)system in *D. melanogaster* (Ganetzky 1999; Kusano et al. 2003). First identified by Hiraizumi (Sandler et al. 1959), SD acts post-meiotically and leads to the reduced condensation and subsequent dysfunction of spermatids in the sperm bundle (Kettaneh and Hartl 1980) that contain large arrays of a repetitive satellite (Kusano et al. 2003; Wu et al. 1988).

Thus, in males heterozygous for SD, upto 99% of the functional sperm contain SD, as opposed to the random Mendelian expectation of 50%.

Why are these differences important? The eventual outcome of female meiotic drive and male meiotic dysfunction may appear to be the same – the increased propagation of the selfish chromosome. But there are significant differences. Perhaps the most important is the fact that female meiotic drive does not entail any drop in fertility, or the number of eggs produced, while male meiotic dysfunction could result in a 50% drop in overall sperm count. In isolation, this fact may not seem profound. After all, most plants and animals make a significantly larger investment in producing eggs as compared to sperm or pollen. Thus, they can "afford to" make a lot more sperm and pollen than they could conceivably need. However, these sperm face stiff competitive threats from individuals that have not been burdened with such a precipitous drop in fertility. If this competition was between the X and Y chromosomes (if the X chromosome were to make the Y dysfunctional, for instance (Jutier et al. 2004)), this would lead to a dramatically skewed sex ratio. Of course, female meiotic drive between Z and W chromosomes (when the female sex is heterogametic) would also lead to alterations of sex ratios.

However, its more benign nature also make female meiotic drive much harder to detect. Maize knobs and gross chromosomal rearrangements (Robertsonian fusions, B chromosomes) are easy to detect cytologically and it is unsurprising that these represented all the known examples of female meiotic drive, until very recently. Conceivably, this kind of meiotic drive could be very common but go undetected for cytologically normal chromosomes in the absence of detailed genotypic data. Recent studies have provided just such genotypic data and confirm that even seemingly normal chromosomes can participate in this selfish battle for evolutionary dominance (Fishman and Willis 2005) and reveal an underlying cost to male meiosis (Fishman and Saunders 2008).

2.6 The Centromere-Drive Model

Taking together the finding that centromeric histones were subject to positive selection as well as the rapid evolution and increased size of centromeric DNA in plants and animals, we proposed an extension of the oötid-competition model (Zwick et al. 1999), which we termed "centromere-drive" (Henikoff et al. 2001; Henikoff and Malik 2002). Under this model, centromeres and centromeric histones evolve under genetic conflict in two steps (Fig. 2.4) (Malik and Bayes 2006). In the first step, an expansion of the centromeric DNA (by recombination) could create a centromere that better attracts microtubules. If this increased microtubule binding conferred an advantage to this centromere expansion in female meiosis, then this would begin sweeping through the population. A number of negative effects can be associated with a sweep of a "selfish centromere," including the fixation of linked deleterious mutations.

These effects would be even more pronounced in the case of the sex chromosomes. For instance, in the case of ZW heterogametic systems (birds, lepidopterans),

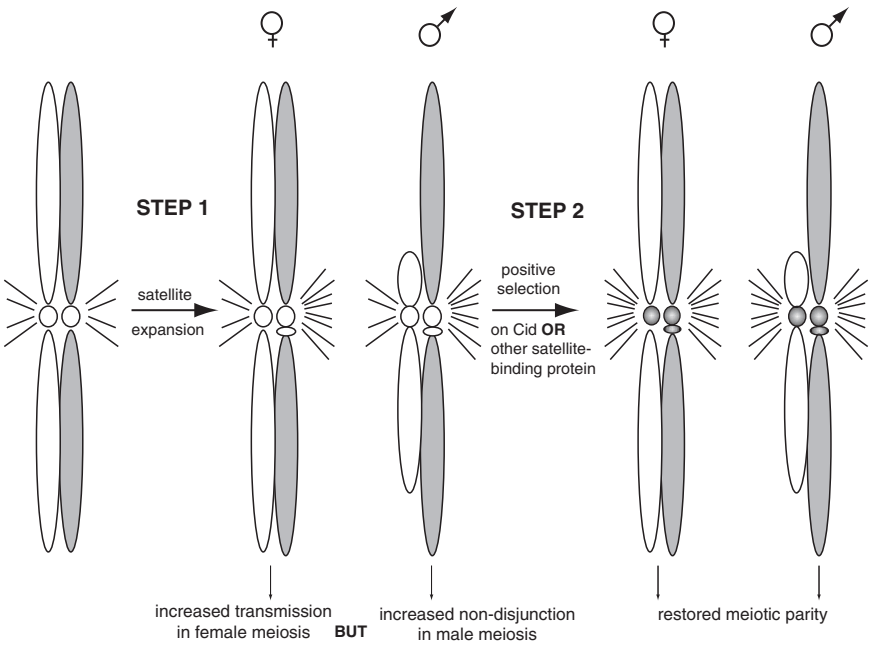


Fig. 2.4 The centromere-drive model (Henikoff et al. 2001; Henikoff and Malik 2002; Malik and Bayes 2006; Malik and Henikoff 2002). In the first stage, a satellite expansion leads to a centromere with enhanced microtubule binding abilities, which can lead to a transmission advantage in female meiosis. This can lead to deleterious effects, including enhanced non-disjunction in male meiosis. In the second stage, a suppressor allele in CenH3 or any other satellite-binding protein that can restore meiotic parity, either by increasing microtubule binding by other centromeres as shown or by reducing microtubule binding by the driving centromere expansion (not shown), will be selectively favored because of its alleviating the deleterious effects of centromere-drive. Thus, genetic conflict between two essential genetic elements can nonetheless drive centromeres to become larger, and CenH3s to be under positive selection

competition between the sex chromosomes for inclusion into the egg would lead to skewed sex ratios and threaten the population. In the case of the XY males (mammals, flies), competition between the X chromosomes would lead to "stronger X centromeres" emerging via selective advantage, but in XY meiosis which relies on symmetry, this would lead to greater nondisjunction, and in extreme instances, sterility (due to recurrent meiotic checkpoint-induced apoptosis) (Eaker et al. 2001; McKee et al. 1998). The situation in the human population where Robertsonian fusions are preferentially transmitted through female meiosis but lead to male sterility is a direct example of just such an effect, and fits all predictions of the "centromere-drive" model (Daniel 2002). A second example has been recently uncovered in monkeyflowers, wherein a strong female meiotic drive has profound consequences on male fertility (Fishman and Saunders 2008; Fishman and Willis 2005).

In such a scenario, any suppressor alleles in autosomal proteins that could alleviate the deleterious effects of this meiotic drive would be selectively swept through this imperiled population. We believe that CenH3s and any heterochromatin binding protein that could restore meiotic parity would serve as such suppressor alleles. For instance, CenH3 is under positive selection to maintain meiotic parity by modulating its DNA-binding preference to deny a satellite expansion the transmission advantage in female meiosis. On the other hand, satellite-binding proteins could restore male meiosis by binding the expanded satellite and preventing CenH3 recruitment; they would also serve as suppressors. Consistent with this prediction, our investigation of satellite-binding proteins and other heterochromatin proteins has revealed that several of them appear to also be evolving under positive selection (J. Bayes and H.S. Malik, unpublished) (Vermaak et al. 2005).

2.7 The "Centromere-Drive" Model is Not Equivalent to the "Molecular-Drive" Model

Since the centromere-drive model was proposed, researchers have often confused it with the proposal of "molecular drive" first coined by Dover (Dover et al. 1982). In fact, these two models are completely dissimilar, with vastly different predictions of the role that selection plays in the process and vastly different trajectories of predicted changes. Since there has been some confusion, we wish to highlight significant differences between the two models, specifically because both are designed to explain the evolutionary dynamics of satellite repeats.

Molecular drive describes evolutionary processes that change the genetic composition of a population through DNA turnover mechanisms. Importantly, molecular drive operates independently of natural selection and genetic drift. Multigene families, in theory, provide the best example of where such process could occur. This is because tandem copies (multigene families), such as those for centromeric DNA satellite repeats, are subject to gene conversion, unequal crossing-over, transposition, slippage replication, and other exchanges. Because mutations changing the sequence of one copy are less common than deletions, duplications, and replacement of one copy by another, the copies gradually come to resemble each other much more than they would if they had been evolving independently. It is important to point out that the process of recombination per se does not increase or even affect the overall probability of mutations being retained in the array (see Sect. 4.2.1). This is because by definition, concerted evolution is unbiased, in which case every version has an equal probability of being the one that replaces the others. However, if the molecular events have any bias favoring one version of the sequence over others, that version will dominate the process and eventually replace the others. The name "molecular drive" reflects the similarity of the process with what was originally the better-known process of meiotic drive. This was intended to affect a biased gene conversion process, which in theory could rapidly accelerate the fixation of mutations in the array. If a protein was to

bind and recognize this array, then under the "molecular" drive model, it would be selected to accommodate the changes that have taken place in the underlying DNA sequence. Several theorists have commented on the population genetic scenarios under which molecular drive might occur, but there are several points to consider when applied to centromeric satellites.

First, research in recombination has shown that the process of biased recombination, as seen in recombination hot spots, is inherently transient because the biased gene conversion actually eliminates the template that was biasing the process (and not the other way around). Even if that were not the case, the fact is that selection is always operating on the satellite DNA sequences. If the sequence were to adopt an unfavorable conformation, for instance, it would perturb centromeric function and be selected against. Thus, the only changes that would be allowed to proliferate would be either neutral changes or those that enhance recruitment of centromeric proteins. Under the neutral scenario, there is no impetus to explain the adaptive evolution of centromeric proteins (essentially, it is the deleterious effects associated with centromeric changes that provides the selective forces that alter the proteins). The model that assumes a "benefit" to the proliferating satellite via biased gene-conversion is consistent with the original model proposed by Dover. Even under this model, if all that was happening was an optimization for the binding of centromeric proteins and DNA, there would be no impetus for the recurrent changes in centromeric proteins. Thus, the rapid evolution of DNA-binding proteins like the CenH3s provides the strongest discriminative features between the models of "molecular-drive" vs. "centromere drive."

Philosophically, the process of "molecular drive" was proposed as a counterpoint to the "selfish gene" theory proposed by several researchers, including Dawkins (Dawkins 1976). In contrast, selfishness is central to the "centromere-drive" model and so is diametrically opposite to the molecular drive model. Nevertheless, the centromere drive model, driven by purely Darwinian means, can fully account for the duality of rapid evolution in both satellite DNA and proteins. In the first instance, satellite DNA changes either in sequence or copy number to enhance binding and subvert meiosis in its own favor, and in the second step, centromeric proteins adapt to suppress the deleterious effects that are concomitant with "selfish" centromeres. This is the form of meiotic drive that was first envisaged by Sandler and Novitski (1957) and is completely explained only in the presence of selection (Burt and Trivers 2006).

2.8 The Centromere-Drive Model in Different Taxonomic Groups

The major driving force in centromere complexity thus appears to be the invention of asymmetric female meiosis. Intriguingly, this invention appears to have happened at least three times independently in the course of plant, algal, and animal evolution. Because of this, we can expect certain predictions about centromere

complexity to hold when viewed through the prism of how meiosis occurs in certain taxa (Henikoff et al. 2001; Malik and Bayes 2006; Malik and Henikoff 2002).

For instance, in fungi, there is largely no differential success based on positioning of meiotic products (note that there may very well be differential success in some instances like filamentous fungi which have linear, rather than tetrahedral asci). Therefore, fungi like *S. pombe* represent the basal state of centromere complexity in eukaryotes, shaped by the presence of both mitosis and meiosis. *S. cerevisiae* represent a further simplification of the centromere configurations represented by *S. pombe*, but this appears to be driven by other factors, including the dramatic loss of the heterochromatin/RNAi machinery that helps to epigenetically define the *S. pombe* centromere (Malik and Henikoff 2009).

Plants and animals, almost all of which possess symmetric (male) and asymmetric (female) meiosis, are subject to episodes of centromere-drive and suppression. As a result, their centromeres are considerably larger than those in *S. pombe*, and consequently, their centromeric proteins are very rapidly evolving, presumably to counteract the deleterious effects of rapid centromere expansions (see Sect. 4.4.1). We expect to see a fairly strict correlation between the presence of male and female meiosis, with both these traits, that is, rapidly evolving centromeric DNA and proteins.

Some interesting deviations from this will be quite instructive. For instance, ciliated protozoans like *Tetrahymena thermophila* have only (asymmetric) female meiosis (Cervantes et al. 2006). Therefore, in this instance, we again expect to see rapid evolution of centromeric DNAs as chromosomes vie for meiotic success. Yet if the deleterious effects are most manifest in male meiosis, it is conceivable that the absence of male meiosis may have obviated the need for centromeric proteins to suppress centromere drive. This may even have motivated the loss of male meiosis in this taxonomic group. Under this scenario, we might predict no positive selection in centromeric proteins. The net result would be unsuppressed centromere drive, leading to greater satellite DNA accumulation. Circumstantial evidence appears to support the idea that centromeres in *T. thermophila* are quite large and they represent the largest fraction of the germline (meiotic) micronuclear genome that is eliminated in the process of forming the (somatic) macronucleus.

Finally, some organisms like bdelloid rotifers appear to have lost meiosis altogether (Mark Welch and Meselson 2000). These could be very instructive to discern the effects of meiosis on the complexity of centromeres in general (loss of heterochromatin, simplicity of centromeres), separate from the instances where asymmetry in female meiosis has evolved. Thus, genetic opportunities afforded to chromosomes to compete with each other during meiosis provide a satisfyingly simple rationale for the bewildering range and rapid evolution of centromeric components that are essential for all forms of chromosome segregation in eukaryotes.

Acknowledgements The author's lab is supported for its studies on centromeres by a grant from the National Institutes of Health (R01-GM74108). The author is grateful to Josh Bayes and Danielle Vermaak for their comments and help with figures.

References

Ahmad K, Henikoff S (2002) Histone H3 variants specify modes of chromatin assembly. Proc Natl Acad Sci USA 99(Suppl 4):16477–16484

Bensasson D, Zarowiecki M, Burt A, Koufopanou V (2008) Rapid evolution of yeast centromeres in the absence of drive. Genetics 178:2161–2167

Black BE, Foltz DR, Chakravarthy S, Luger K, Woods VL Jr, Cleveland DW (2004) Structural determinants for generating centromeric chromatin. Nature 430:578–582

Black BE, Jansen LE, Maddox PS, Foltz DR, Desai AB, Shah JV, Cleveland DW (2007) Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol Cell 25:309–322

Blower MD, Karpen GH (2001) The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat Cell Biol 3:730–739

Blower MD, Sullivan BA, Karpen GH (2002) Conserved organization of centromeric chromatin in flies and humans. Dev Cell 2:319–330

Buchwitz BJ, Ahmad K, Moore LL, Roth MB, Henikoff S (1999) A histone-H3-like protein in *C. elegans*. Nature 401:547–548

Buckler EST, Phelps-Durr TL, Buckler CS, Dawe RK, Doebley JF, Holtsford TP (1999) Meiotic drive of chromosomal knobs reshaped the maize genome. Genetics 153:415–426

Burt A, Trivers R (2006) Genes in conflict. Belknap Press

Cervantes MD, Xi X, Vermaak D, Yao MC, Malik HS (2006) The CNA1 histone of the ciliate Tetrahymena thermophila is essential for chromosome segregation in the germline micronucleus. Mol Biol Cell 17:485–497

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

Clarke L, Baum MP (1990) Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences. Mol Cell Biol 10:1863–1872

Cooper JL, Henikoff S (2004) Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol 21:1712–1718

Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD, McCombie WR, Martienssen RA, Marra M, Preuss D (1999) Genetic definition and sequence analysis of Arabidopsis centromeres. Science 286:2468–2474

Daniel A (2002) Distortion of female meiotic segregation and reduced male fertility in human Robertsonian translocations: consistent with the centromere model of co-evolving centromere DNA/centromeric histone (CENP-A). Am J Med Genet 111:450–452

Dawe RK, Cande WZ (1996) Induction of centromeric activity in maize by suppressor of meiotic drive 1. Proc Natl Acad Sci USA 93:8512–8517

Dawkins R (1976) The selfish gene. Oxford University Press

Dover GA, Strachan T, Coen ES, Brown SD (1982) Molecular drive. Science 218:1069

Eaker S, Pyle A, Cobb J, Handel MA (2001) Evidence for meiotic spindle checkpoint from analysis of spermatocytes from Robertsonian-chromosome heterozygous mice. J Cell Sci 114:2953–2965

Fishman L, Saunders A (2008) Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. Science 322:1559–1562

Fishman L, Willis JH (2005) A novel meiotic drive locus almost completely distorts segregation in mimulus (monkeyflower) hybrids. Genetics 169:347–353

Fitzgerald-Hayes M, Clarke L, Carbon J (1982) Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. Cell 29:235–244

Ganetzky B (1999) Yuichiro Hiraizumi and forty years of segregation distortion. Genetics 152:1–4

Haaf T, Willard HF (1997) Chromosome-specific alpha-satellite DNA from the centromere of chimpanzee chromosome 4. Chromosoma 106:226–232

Halverson D, Baum M, Stryker J, Carbon J, Clarke L (1997) A centromere DNA-binding protein from fission yeast affects chromosome segregation and has homology to human CENP-B. J Cell Biol 136:487–500

Halverson D, Gutkin G, Clarke L (2000) A novel member of the Swi6p family of fission yeast chromo domain-containing proteins associates with the centromere in vivo and affects chromosome segregation. Mol Gen Genet 264:492–505

Hartl DL (1975) Modifier theory and meiotic drive. Theor Popul Biol 7:168–174

Hawley RS, Irick H, Zitron AE, Haddox DA, Lohe A, New C, Whitley MD, Arbel T, Jang J, McKim K (1992). There are two mechanisms of achiasmate segregation in Drosophila females, one of which requires heterochromatic homology. Dev Genet 13:440–467

Henikoff S, Malik HS (2002) Centromeres: selfish drivers. Nature 417:227

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098–1102

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Jutier D, Derome N, Montchamp-Moreau C (2004) The sex-ratio trait and its evolution in Drosophila simulans: a comparative approach. Genetica 120:87–99

Karpen GH, Le MH, Le H (1996) Centric heterochromatin and the efficiency of achiasmate disjunction in Drosophila female meiosis. Science 273:118–122

Kettaneh NP, Hartl DL (1980) Ultrastructural analysis of spermiogenesis in segregation distorter males of Drosophila melanogaster: the homozygotes. Genetics 96:665–683

Kusano A, Staber C, Chan HY, Ganetzky B (2003) Closing the (Ran)GAP on segregation distortion in Drosophila. Bioessays 25:108–115

Lee HR, Zhang W, Langdon T, Jin W, Yan H, Cheng Z, Jiang J (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. Proc Natl Acad Sci USA 102:11793–11798

Lo AW, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E, Magliano DJ, Choo KH (2001) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. EMBO J 20:2087–2096

Lohe AR, Brutlag DL (1987) Identical satellite DNA sequences in sibling species of Drosophila. J Mol Biol 194:161–170

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389:251–260

Malik HS, Bayes JJ (2006) Genetic conflicts during meiosis and the evolutionary origins of centromere complexity. Biochem Soc Trans 34:569–573

Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics 157:1293–1298

Malik HS, Henikoff S (2002) Conflict begets complexity: the evolution of centromeres. Curr Opin Genet Dev 12:711–718

Malik HS, Henikoff S (2003) Phylogenomics of the nucleosome. Nat Struct Biol 10:882–891

Malik HS, Henikoff S (2009) Major evolutionary transitions in centromere complexity. Cell (submitted)

Mark Welch D, Meselson M (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science 288:1211–1215

McAllister BF, Werren JH (1999) Evolution of tandemly repeated sequences: What happens at the end of an array? J Mol Evol 48:469–481

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654

McKee BD, Wilhelm K, Merrill C, Ren X (1998) Male sterility and meiotic drive associated with sex chromosome rearrangements in Drosophila. Role of X-Y pairing. Genetics 149:143–155

Merrill C, Bayraktaroglu L, Kusano A, Ganetzky B (1999) Truncated RanGAP encoded by the segregation distorter locus of Drosophila. Science 283:1742–1745

Monen J, Maddox PS, Hyndman F, Oegema K, Desai A (2005) Differential role of CENP-A in the segregation of holocentric C. elegans chromosomes during meiosis and mitosis. Nat Cell Biol 7:1248–1255

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145
Nielsen J, Wohlert M (1991) Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. Hum Genet 87:81–83
Palmer DK, O'Day K, Wener MH, Andrews BS, Margolis RL (1987) A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J Cell Biol 104:805–815
Pardo-Manuel de Villena F, Sapienza C (2001a) Female meiosis drives karyotypic evolution in mammals. Genetics 159:1179–1189
Pardo-Manuel de Villena F, Sapienza C (2001b) Nonrandom segregation during meiosis: the unfairness of females. Mamm Genome 12:331–339
Platero JS, Ahmad K, Henikoff S (1999) A distal heterochromatic block displays centromeric activity when detached from a natural centromere. Mol Cell 4:995–1004
Rhoades M (1942) Preferential segregation in maize. Genetics 27:395–407
Rudd MK, Wray GA, Willard HF (2006) The evolutionary dynamics of alpha-satellite. Genome Res 16:88–96
Samonte RV, Ramesh KH, Verma RS (1997) Comparative mapping of human alphoid satellite DNA repeat sequences in the great apes. Genetica 101:97–104
Sandler L, Novitski E (1957) Meiotic drive as an evolutionary force. Am Nat 41:105–110
Sandler L, Hiraizumi Y, Sandler I (1959) Meiotic drive in natural populations of Drosophila melanogaster. I. The cytogenetic basis of segregation-distortion. Genetics 44:233–250
Sawamura K, Yamamoto MT (1993) Cytogenetical localization of Zygotic hybrid rescue (Zhr), a Drosophila melanogaster gene that rescues interspecific hybrids from embryonic lethality. Mol Gen Genet 239:441–449
Sawamura K, Yamamoto MT, Watanabe TK (1993) Hybrid lethal systems in the *Drosophila melanogaster* species complex. II. The Zygotic hybrid rescue (Zhr) gene of *D. melanogaster*. Genetics 133:307–313
Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. Science 294:109–115
Shelby RD, Vafa O, Sullivan KF (1997) Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. J Cell Biol 136:501–513
Stephan W (1989) Tandem-repetitive noncoding DNA: forms and forces. Mol Biol Evol 6:198–212
Stephan W, Cho S (1994) Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. Genetics 136:333–341
Stoler S, Keith KC, Curnick KE, Fitzgerald-Hayes M (1995) A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev 9:573–586
Sullivan KF, Hechenberger M, Masri K (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol 127:581–592
Sun X, Le HD, Wahlstrom JM, Karpen GH (2003) Sequence analysis of a functional Drosophila centromere. Genome Res 13:182–194
Takahashi K, Chen ES, Yanagida M (2000) Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. Science 288:2215–2219
Talbert PB, Bryson TD, Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. J Biol 3:18
Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. Plant Cell 14:1053–1066
Tao Y, Hartl DL, Laurie CC (2001) Sex-ratio segregation distortion associated with reproductive isolation in Drosophila. Proc Natl Acad Sci USA 98:13183–13188
Vermaak D, Hayden HS, Henikoff S (2002) Centromere targeting element within the histone fold domain of Cid. Mol Cell Biol 22:7553–7561
Vermaak D, Henikoff S, Malik HS (2005) Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. PLoS Genet 1:96–108
Walsh JB (1987) Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics 115:553–567

Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, Davis P, Feltwell T, Fraser A, Gentles S, Goble A, Hamlin N, Harris D, Hidalgo J, Hodgson G, Holroyd S, Hornsby T, Howarth S, Huckle EJ, Hunt S, Jagels K, James K, Jones L, Jones M, Leather S, McDonald S, McLean J, Mooney P, Moule S, Mungall K, Murphy L, Niblett D, Odell C, Oliver K, O'Neil S, Pearson D, Quail MA, Rabbinowitsch E, Rutherford K, Rutter S, Saunders D, Seeger K, Sharp S, Skelton J, Simmonds M, Squares R, Squares S, Stevens K, Taylor K, Taylor RG, Tivey A, Walsh S, Warren T, Whitehead S, Woodward J, Volckaert G, Aert R, Robben J, Grymonprez B, Weltjens I, Vanstreels E, Rieger M, Schafer M, Muller-Auer S, Gabel C, Fuchs M, Dusterhoft A, Fritzc C, Holzer E, Moestl D, Hilbert H, Borzym K, Langer I, Beck A, Lehrach H, Reinhardt R, Pohl TM, Eger P, Zimmermann W, Wedler H, Wambutt R, Purnelle B, Goffeau A, Cadieu E, Dreano S, Gloux S, Lelaure V, Mottier S, Galibert F, Aves SJ, Xiang Z, Hunt C, Moore K, Hurst SM, Lucas M, Rochet M, Gaillardin C, Tallada VA, Garzon A, Thode G, Daga RR, Cruzado L, Jimenez J, Sanchez M, del Rey F, Benito J, Dominguez A, Revuelta JL, Moreno S, Armstrong J, Forsburg SL, Cerutti L, Lowe T, McCombie WR, Paulsen I, Potashkin J, Shpakovski GV, Ussery D, Barrell BG, Nurse P (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature 415:871–880

Wu CI, Lyttle TW, Wu ML, Lin GF (1988) Association between a satellite DNA sequence and the responder of segregation distorter in *D. melanogaster*. Cell 54:179–189

Wu CI, True JR, Johnson N (1989) Fitness reduction associated with the deletion of a satellite DNA array. Nature 341:248–251

Yoda K, Ando S, Morishita S, Houmura K, Hashimoto K, Takeyasu K, Okazaki T (2000) Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. Proc Natl Acad Sci USA 97:7266–7271

Yoda K, Morishita S, Hashimoto K (2004) Histone variant CENP-A purification, nucleosome reconstitution. Methods Enzymol 375:253–269

Zhang P, Knowles BA, Goldstein LS, Hawley RS (1990) A kinesin-like protein required for distributive chromosome segregation in Drosophila. Cell 62:1053–1062

Zwick ME, Salstrom JL, Langley CH (1999) Genetic variation in rates of nondisjunction: association of two naturally occurring polymorphisms in the chromokinesin nod with increased rates of nondisjunction in *Drosophila melanogaster*. Genetics 152:1605–1614

Chapter 3
Centromere-Competent DNA: Structure and Evolution

Đurđica Ugarković

Contents

Abstract Although extant data favour centromere being an epigenetic structure, it is also clear that centromere formation is based on DNA, in particular, tandemly repeated satellite DNA and its transcripts. Presence of conserved structural motifs within satellite DNAs such as periodically distributed AT tracts, protein binding sites, or promoter elements indicate that despite sequence flexibility, there are structural determinants that are prerequisite for centromere function. In addition, existence of functional centromeric DNA transcripts indicates possible importance of structural elements at the level of RNA secondary or tertiary structure. Rapid centromere evolution is explained by homologous recombination followed by extrachromosomal rolling circle replication. This could lead to amplification of different satellite sequences within a genome. However, only those satellites that have inherent

Đ. Ugarković
Department of Molecular Biology, Ruđer Bošković Institute, Bijenička 54,
HR-10002, Zagreb, Croatia
e-mail: ugarkov@irb.hr

Đ. Ugarković (ed.), *Centromere,* Progress in Molecular and Subcellular Biology 48,
DOI: 10.1007/978-3-642-00182-6_3, © Springer-Verlag Berlin Heidelbarg 2009

centromere-competence in the form of structural requirements necessary for centromere function are after amplification fixed in a population as a new centromere.

3.1 Introduction

The centromere is a region of the chromosome that enables the accurate partition of newly replicated sister chromatids between daughter cells during mitosis and meiosis. It holds sister chromatids together and through its centromere DNA–protein complex known as the kinetochore binds spindle microtubules to bring about accurate chromosome movements (Dobie et al. 1999). In addition, centromere regulates progression of cell cycle and is critical in sensing completion of metaphase and triggers the onset of anaphase (Nasmyth 2002). It is visible as the primary constriction on metaphase chromosome.

Centromeric DNA sequences and proteins have been characterized in different organisms, ranging from yeast to human. While a number of proteins shares homology among evolutionarily distant organisms, centromeric DNA sequences differ significantly even among closely related species and evolve rapidly during speciation (Malik and Henikoff 2002). The lack of conservation of centromere DNA could be the characteristic of a single organism as illustrated by neocentromere formation from different genomic sequences in humans (Marshall et al. 2008). Formation of a neocentromere occurs as a result of chromosomal rearrangement that leads to the loss of normal centromere. Most neocentromeres, however, share no sequence homology to normal centromere. Such a plasticity of centromeric DNA could be explained by epigenetic control of centromere function, which does not depend absolutely on primary DNA sequence (Dawe and Henikoff 2006). According to such concept, centromere activation or inactivation might be caused by modifications of chromatin. Such acquired chromatin epigenetic modifications are then inherited from one cell division to the next. Concerning centromere-specific chromatin modification, it is now evident that all centromeres contain a centromere specific histone H3 variant, CenH3, which replaces histone H3 in centromeric nucleosomes and provides a structural basis that differentiates the centromere from the surrounding chromatin. This modified histone H3 is known under different names such as CENP-A (humans), Cid (*Drosophila melanogaster*), or Cse4 (*Saccharomyces cerevisiae*) (reviewed in Black and Bassett 2008; see Chap. 1 in this book). CenH3 is characteristic not only for normal centromeres but also for neocentromeres and is essential for the establishment and maintenance of centromere function. Centromeric nucleosomes are specific not only by the presence of CenH3, but also by their internal organization. They seem to be organized as a tetramer composed of one molecule each of CenH3, H2A, H2B, and H4, different from the octamer found in bulk nucleosomes (Dalal et al. 2007). CenH3 chromatin is localized in the inner kinetochore plate and it seems that it exhibits greater conformational rigidity necessary to maintain the architecture during metaphase when tension pulls the kinetochore towards the poles (reviewed in Vagnarelli et al. 2008).

3.2 Types of Centromere

Although extant data favour centromere being an epigenetic structure, it is also clear that centromere formation is based on DNA, and as new results suggest, also very probably on RNA. A most simple centromere characteristic for budding yeast *S. cerevisiae* is referred as a point centromere, as it encompasses a short distinct DNA sequence of approximately 125 bp, which contains no repetitive DNA. This sequence specifies a kinetochore formation and such simple centromere binds a single microtubule (Kalitsis 2008). More complex, regional centromeres are common for higher eukaryotes, including fission yeast *Schizosaccharomyces pombe*. They encompass longer, usually Mb size arrays composed of repetitive sequences and form a larger kinetochore that interacts with a number of microtubules. The common feature of regional centromeres across the wide species range, which includes *Arabidopsis thaliana*, rice, maize, *D. melanogaster*, and humans, is the presence of satellite DNA as their predominant component (Schueler et al. 2001; Kumekawa et al. 2001; Sun et al. 2003; Jin et al. 2004; Zhang et al. 2004). In the case of human chromosomes, the main centromeric component is alpha satellite DNA. Human alpha satellite DNA makes up 3–5% of each chromosome and the fundamental repeat unit is based on diverged 171 bp monomers. Monomers are tandemly arranged into long homogenous arrays of 250 kb to more than 4 Mb per chromosome (Ugarković 2008a). Alpha satellite DNA is not absolutely necessary for centromere formation, because in its absence euchromatic DNA is capable of being activated to form a neocentromere (Amor and Choo 2002). However, studies of de novo chromosome formation have revealed the preferential formation of centromere on stretches composed of tandemly repeated satellite DNA (Grimes et al. 2002). For example, de novo assembly of human centromere occurs on alpha satellite DNA array, which contains a 17 bp binding motif for centromeric protein B (CENP-B) known as CENP-B box (Grimes et al. 2002; Masumoto et al. 2004). The studies show that alpha satellite is a preferred substrate for centromere formation and that CENP-B box plays an essential role in centromere establishment. However, once established, centromere seems to be further propagated and maintained without CENP-B protein (Okada et al. 2007).

These examples reveal that point centromeres are restricted completely to particular DNA sequence, while in regional centromere this restriction is a partial one. On the other hand, there are examples when centromeres are not localized to any particular chromosomal region. Such diffuse centromeres of holocentric chromosomes of nematodes are distributed along the lengths of the chromosomes attaching to microtubules at many sites (Maddox et al. 2004). The character of DNA sequences that are responsible for the establishment of diffuse centromeres is not defined. However, sequencing of genome of nematode *Caenorhabditis elegans* revealed the presence of many families of short interspersed repeats. Some of them, after cloning into suitable vectors and introduction into yeast *S. cerevisiae* are shown to contribute to increased mitotic stability of plasmids, indicative of centromeric role (Kalitsis 2008).

In addition to DNA and proteins, RNA seems also to be a structural component of centromere. Transcripts of alpha satellite DNAs have been shown to be a functional component of the kinetochore, participating in recruitment of kinetochore proteins (Wong et al. 2007). In addition, ribonucleoprotein complexes are required for mitotic spindle assembly (Blower et al. 2005). All these data point to an important role for DNA and RNA, in particular, tandemly repeated satellite DNA and its transcripts in centromere/kinetochore establishment and function. New findings related to evolutionary constraints on centromeric satellite DNAs also shed more light on the possible role of these sequences. Despite sequence heterogeneity among species, the common pattern of DNA structural motifs required for centromere specification is beginning to be discerned.

3.3 Evolutionary Mechanisms Affecting Centromeric DNA

3.3.1 Role of Stochastic Processes

In general, centromeric regions are considered the most rapidly evolving compartments in the eukaryotic genome. In the case of point centromere, high mutation rate seems to be responsible for such a rapid sequence change (Bensasson et al. 2008). Regional centromeres, however, which are characterized by repetitive structure, mostly in the form of tandem satellite DNA repeats exhibit change not only in sequence but also in repeat copy number. Therefore, evolution of regional centromere proceeds not only by mutations but also by recombination. Recombinational mechanisms such as gene conversion and unequal crossingover affect repetitive DNAs and are responsible for the rapid horizontal spread of newly occurring mutations among monomers within a repetitive family. This results ultimately in homogenization of changes among repeats within the genome and their subsequent fixation in members of reproductive populations in a process known as molecular drive (Dover 1986). This mode of horizontal evolution, characteristic for repetitive families, is known as concerted evolution. The process of homogenization occurs at species-specific rates but is faster and independent of the mutation rate. As a result of concerted evolution, repeats of a satellite DNA within regional centromere exhibit high homology within a species. However, because of the same process, different mutations are randomly fixed in reproductively isolated populations, causing rapid divergence of centromere sequence among species.

Besides being responsible for the spreading of mutations horizontally through members of the repetitive family, unequal crossingover is also responsible for changes in repetitive DNA copy number, affecting in this way the length of centromere arrays (Smith 1976). Theoretical studies on satellite DNA dynamics explain its loss from the genome by unequal crossingover, demonstrating an inverse correlation between the rate of unequal crossingover and the preservation time of the satellite DNA (Stephan 1986). Satellite DNAs can also increase in copy number either by

replication slippage, rolling circle replication, and conversion-like mechanisms in a relatively short evolutionary time (reviewed in Ugarković and Plohl 2002). The outcome of all these mechanisms affecting satellite DNA arrays is a high turnover of centromeric and pericentromeric regions of the eukaryotic genome. On the model of mouse cells, it has been shown that centromere mitotic recombination occurs at a much higher frequency than chromosome arm recombination, and is controlled by the epigenetic state of centromeric heterochromatin, in particular by centromeric DNA methylation (Jaco et al. 2008). Methylation of centromeric DNA represses illicit recombination at repeated satellite DNA and is suggested to be important for the maintenance of centromere integrity. On the other hand, the reduced frequency of recombination in the neighborhood of centromeres during meiosis, relative to the rest of chromosome, has been documented in *D. melanogaster* and many other organisms (Charlesworth et al. 1986; Stephan 2007). It has been proposed that the reduced meiotic recombination could be the consequence of natural selection, which lowers the unequal exchange between repeats and in this way prevents significant change in repetitive array lengths. Repeat length change could lead to the variation in the number of microtubule binding sites per chromosome, which can further result in nondisjunction events and aneuploidy.

3.3.2 Role of Natural Selection

In addition to stochastic, random processes that affect centromeric DNA and induce its rapid sequence evolution, there are indications for the natural selection shaping evolution of centromeric DNA sequence (Ugarković 2005). This indication is based on the extreme sequence preservation and wide evolutionary distribution of some satellite DNAs as well as on the conservation of particular structural motifs. Selection was first thought to influence satellite DNA sequences following the observation of nonrandom distribution of variability along the satellite monomers, resulting in constant and variable regions in *Arabidopsis thaliana* and human alpha satellite DNA (Romanova et al. 1996; Heslop-Harrison et al. 1999). Nonrandom pattern of variability was subsequently detected in many centromeric satellites (Hall et al. 2003; Mravinac et al. 2004; 2005), as well as preservation of variability at particular positions within a satellite in different populations (Feliciello et al. 2005). Restricted variability could be probably related to interaction of satellite DNAs with specific proteins necessary for heterochromatin and centromere formation as well as to the role of satellite DNAs in controlling gene expression.

The best characterized satellite DNA-binding protein is human centromere protein B (CENP-B), which binds to a 17 bp motif in human alpha satellite DNA known as the CENP-B box (Masumoto et al. 1989). Proteins homologous to CENP-B have been found in many eukaryotes, including the fission yeast *S. pombe*, and motifs that are 60–70% similar to the CENP-B box have been detected in diverse centromeric repeats of mammals and insects (Kipling and Warburton 1997; Mravinac et al. 2004; Fig. 3.1). Although only 23% of repeats in human α satellite DNA have a functional CENP-B

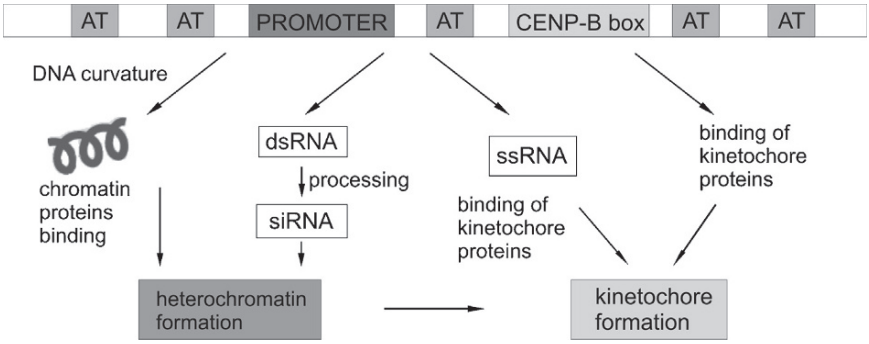


Fig. 3.1 Evolutionary constraints on centromeric satellite DNAs. Structural requirements posed on satellite DNAs which enable them to be retained in the genome as members of satellite library and to be potentially expanded into a "new" centromere might include periodic clusters of A + Ts, binding sites for centromeric proteins such as CENP-B box, or promoter elements necessary for active transcription. Periodic distribution of AT tracts leads to curvature of the DNA helix axis and formation of superhelical tertiary structure thought to be important for heterochromatin establishment. Transcription of satellite DNAs proceeds in the form of either double-stranded RNA (dsRNA) or single-stranded RNA (ssRNA). Long ssRNAs are required for the association of kinetochore proteins, while dsRNA is processed into small interfering RNAs (siRNAs) that participate in heterochromatin formation. Constraints on satellite RNA secondary and/or tertiary structure could exist in order to preserve its ability to bind kinetochore proteins

box, it seems to be essential for the assembly of centromere-specific chromatin and centromere establishment, but not for the centromere maintenance (Ohzeki et al. 2002; Basu et al. 2005; Okada et al. 2007).

Satellite DNAs are usually AT rich but A's or T's are not randomly distributed within the sequence. Clustering of A or T and regular phasing of A or T ≥3 tracts has been reported for many different satellite DNAs, including human alpha satellite DNA (Martinez-Balbas et al. 1990; Ugarković et al. 1996a; Fig. 3.1). Periodic distribution of AT tracts usually induces curvature of the DNA helix axis and formation of tertiary structure in the form of a superhelix (Fitzgerald et al. 1994). Such a structure is thought to be important for the tight packing of DNA and proteins in heterochromatin (Ugarković et al. 1992).

Palindromic sequences that could potentially lead to the formation of dyad structures are common elements of centromeric and pericentromeric satellite DNAs in budding yeast, insects, and human (Tal et al. 1994; Ugarković et al. 1996b; Zhu et al. 1996). It is not clear if they perform some function, but it can be hypothesized that some palindromic sequences could be recognized by DNA binding proteins, such as transcription factors. Some homeodomain proteins like Pax3, which is known to play an important role during neurogenesis, bind short palindromes present within major mouse satellite DNA (personal communication). The recent investigation has revealed that the topoisomerase II recognizes and cleaves a specific hairpin structure formed by alpha satellite DNA (Jonstrup et al. 2008). It has been

suggested that a subpopulation of the cellular topoisomerase II located at centromeres plays a role for sister chromatid cohesion in the centromeric region. The hairpin cleavage therefore could be connected to a cohesion role of topoisomerase II at centromeres.

Other functional motifs and regulatory elements for RNA polymerase (pol) II and RNA pol III are predicted in some satellite sequences (Renault et al. 1999; Fig. 3.1). Human satellite III, which is specifically expressed under stress, has a binding motif for the heat shock transcription factor 1 that drives RNA pol II transcription (Metz et al. 2004). In schistosome satellite DNA, which encodes an active ribozyme, a functional RNA pol III promoter is present (Ferbeyre et al. 1998). The sequence of satellite 2 found in the newts *Notophthalmus viridescens* and *Triturus vulgaris meridionalis* contains a functional analogue of the vertebrate small nuclear RNA (snRNA) promoter that is responsible for RNA pol II transcription (Coats et al. 1994). Promoters for RNA Pol II are also the characteristic of centromeric satellite DNAs from beetle species *Palorus ratzeburgii* and *Palorus subdepressus* (Pezer and Ugarković 2008a; 2009). In general, the presence of functional elements within centromeric satellite DNA sequences points to the role of natural selection in preserving such motifs.

Some centromeric satellites, however, exhibit sequence conservation of the whole monomer sequence for long evolutionary periods. Extreme sequence conservation of two satellite DNAs that represent major pericentromeric repeats in the coleopteran insect species *Palorus ratzeburgii* and *Palorus subdepressus* has been reported (Mravinac et al. 2002; 2005). These satellites are present in many coleopteran species at a low copy number and their sequences have remained unchanged for 60 million years. This remarkable antiquity and sequence conservation are also characteristic of human alpha satellite DNA, which has been detected as a rare, highly conserved repeat in evolutionary distant species such as chicken and zebrafish (Li and Kirby 2003). This complete sequence conservation and the wide evolutionary distribution of some satellite sequences has led to the assumption that, in addition to participating in centromere formation, they could perform some other role possibly acting as *cis*-regulatory elements of gene expression.

In addition to relatively conserved regions found in diverse centromeric satellites, other more variable regions also exist. Variable regions might also be functionally important owing to their interaction with rapidly evolving proteins. Such an example is the centromere-specific histone, CenH3, which replaces histone H3 in centromeric nucleosomes and is required for proper chromosome distribution during cell division (Henikoff and Dalal 2005). Unlike the highly conserved histone H3, CenH3 is divergent and subject to the influence of positive selection, which particularly affects the sites that potentially interact with satellite DNA (Cooper and Henikoff 2004; see Chap. 2 in this book, Sect. 3.2.2). It has been proposed that variable regions within satellite DNA sequence drive the adaptive evolution of specific centromeric histones. In addition to CenH3, other kinetochore proteins exhibit rapid sequence evolution in fly *D. melanogaster* as well as in worm *C. elegans*, while in mammals, plants, and fungi the rate of evolution is much lower (Meraldi et al. 2006).

3.4 Point Centromere DNA and Its Evolution

While in most animals and plants species, centromeres are complex and regional, encompassing long Mb size arrays of highly repetitive, satellite DNA, centromeres in *Saccharomyces* yeast and several other budding yeasts such as *Candida glabrate* and *Kluyvermyces lactis* occupy a very small region of approximately 120 bp and are referred to as point centromere. The centromeric sequence contains no repetitive DNA and consists of three functionally distinct regions: CDEI and CDEIII, which are 8 bp and approximately 25 bp long, respectively, and represent protein binding sites, as well as of CDEII, approximately 90 bp long, which binds centromere-specific histone Cse4 (Hegemann and Fleig 1993). CDEI and CDEIII elements exhibit sequence conservation among different budding yeast species. Mutations in CDEI impair but do not abolish function in mitosis and meiosis, while single base change or short deletions within CDEIII completely inactivate the centromere. CDEII from different chromosomes within same species are highly divergent, up to 60%, but functionally interchangeable (Clarke and Carbon 1983), suggesting that binding of Cse4 is not sequence specific. However, changes in AT content, which is averaging 90%, pattern of homopolymer runs of A's and T's, and length can disrupt centromere function (Baker and Rogers 2005). This indicates that DNA curvature or flexibility which depends on the pattern of distribution of A and T tracts could be related to centromere function. It has been shown that bent and unbent CDEII DNAs, differing at only six nucleotides, displayed a 60-fold difference in mitotic chromosome loss rates. Since AT rich sequences that exhibit homopolymer bias such as CDEII are found predominantly at centromeres of various species, this seems to represent a type of «code» that partially can explain centromere identity.

Periodic distribution of A and T tracts represents a commonality between point *Saccharomyces* centromere and complex regional centromeres of higher organisms. Survey of more than hundred different satellite DNAs revealed that approximately 50% of them exhibit DNA curvature induced by periodic distribution of A or T tracts (Fitzgerald et al. 1994). Such highly nonrandom patterns of A's and T's characterized by homopolymer runs of 5–7 nucleotides might imply influence of selection to preserve mitotic centromere function in *Saccharomyces* as well as in many higher eukaryotes (Baker and Rogers 2005).

Comparison of near-complete sequences of chromosome III from three closely related lineages of the wild yeast *Saccharomyces paradoxus*, which is a relative of S. *cerevisiae*, has shown that the centromere region CDEII is the most rapidly evolving part of the chromosome (Bensasson et al. 2008). This centromere region is evolving faster than sequences that are not under selective constraint. Such rapid evolution could result from elevated mutation rate or influence of positive selection. It has been proposed that positive selection drives rapid fixation of mutations in centromeric regions by imposing a bias in favour of retaining mutations. The positive selection might be due to the advantage conferred to mutated centromere during female meiosis known as «centromere drive hypothesis» (Malik and Henikoff 2002; see Chap. 2 in this book). However, in the case of point *Saccharomyces* centromere,

it seems that elevated mutation rate within CDEII is responsible for the rapid evolution and not positive selection. What on the other hand could induce such a high substitution rate in the yeast centromere region is not clear.

While elevated mutation rate is considered as a major contributor to rapid evolution of point centromere, recombinational mechanisms such as unequal crossing over and gene conversion that preferentially affect segments of repetitive DNA are major genetic mechanisms governing evolution of complex regional centromeres (Ugarković and Plohl 2002). Comprehensive phylogenetic and structural analysis of centromere/kinetochore proteins from different species revealed that organisms with regional and point centromeres have a common ancestor, a fungus containing a regional centromere, implying that simple, point centromere arose from complex, regional centromere (Meraldi et al. 2006).

Different from the regional centromeres that generally have no transcribed genes in their vicinity, transcribed genes are found very close to point *S. cerevisiae* centromeres (Westermann et al. 2007). It is, however, not known if transcripts are structural component of point *Saccharomyces* centromere, as found for complex regional centromeres (Wong et al. 2007).

3.5 Regional Centromere DNA and Its Evolution

Regional centromere encompasses from 1 kb in budding yeast *Candida albicans* (Sanyal et al. 2004) to few megabases in human (Schueler et al. 2001), and is typically composed of repetitive DNA elements, mostly in the form of tandemly repeated satellite DNAs. A single satellite DNA can predominate at the centromeric regions such as the case of alpha satellite DNA at human centromeres (Schueler et al. 2001). In *D. melanogaster* and beetle species *Tribolium madens*, two or more different satellites are interspersed within centromeric regions (Durajlija-Žinić et al. 2000; Sun et al. 2003).

Different centromeric satellite DNAs may persist in the genome usually at centromeric or pericentromeric locations for long evolutionary time forming a collection or library of satellite sequences shared among related lineages (Fry and Salser 1977). The amount of satellite DNAs in a single centromere can be increased or reduced dramatically in a short time frame. Such rapid turnover characteristic for regional centromere evolution can be explained by differential amplification or expansion of satellite DNAs from the library in any species (Ugarković and Plohl 2002). The first experimental demonstration of a satellite DNA library is found in the insect genus *Palorus* (Coleoptera), where all examined species posses a common collection of centromeric satellite DNAs (Meštrović et al. 1998). A different single satellite is significantly amplified or expanded in each of the different species, resulting in species-specific satellite DNA profiles. The existence of satellite libraries is supported for different groups of species, including plants, nematodes, insects, and mammals, as well as their preferential localization within pericentromeric and centromeric regions (King et al. 1995; Vershinin et al. 1996; Cesari et al. 2003; Lin and Li 2006;

Meštrović et al. 2006; Bruvo-Mađarić et al. 2007; Kawabe and Charlesworth 2007). In the marsupial genus *Macropus*, three satellite DNAs are involved in the creation of centromeric arrays in nine examined species (Bulazel et al. 2007; see Chap. 4 in this book). Each species, however, has experienced different expansion and contraction of individual satellites. In *Bovini*, six related centromeric satellite DNAs are shared among species fluctuating considerably in relative amounts (Nijman and Lenstra 2001).

3.5.1 Human Centromeric DNA

Different satellite DNAs that coexist in the same species can vary significantly in their sequence homogeneity and are considered as independent evolutionary units. In addition, each satellite DNA can exist in the form of different, usually chromosome-specific satellite subfamilies (reviewed in Ugarković and Plohl 2002). All primate species share alpha satellite DNA, which in the form of different subfamilies represents the major component of all centromeres (Lee et al. 1997). Alpha satellite is composed of two basic types of repeat units: a 171 bp monomer and higher order repeats (HOR). Higher order repeats have complex repeat units composed of up to 30 diverged 171 bp monomers (Alexandrov et al. 2001) and are characteristic of centromeres of higher primates, while in the genomes of lower primates, monomeric alpha satellite repeats prevail and comprise long centromeric arrays.

The centromeric region has been characterized in detail for the human X chromosome (Fig. 3.2; Schueler et al. 2001). Two evolutionarily distinct classes of alpha satellite are present within the centromeric region of the X chromosome. One class encompasses an approximately 3 Mb array of alpha satellite DNA known as DXZ1, which is present at the primary constriction and is X chromosome specific. This region is defined by a 2.0 kb higher-order repeat, which consists of twelve 171 bp monomers. The canonical higher order repeats are highly homogenous, showing

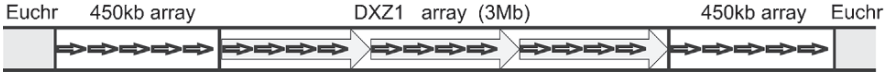


Fig. 3.2 Organization of alpha satellite DNA within centromere of human X-chromosome based on data from Schueler et al. (2005). DXZ1 region of 3 Mb in which primary constriction is located is composed of tandemly repeated higher order repeats (HORs). HORs are mutually highly homologous exhibiting 1–2% divergence. DXZ1 array is flanked on both sites by region of approximate size of 450 kb, which is composed mostly of alpha satellite monomers. Alpha satellite monomers within 450 kb array exhibit divergence between 20% and 30% and are interspersed with transposable elements such as LINE and SINE. Higher order repeats participate in kinetochore formation while diverged monomers contribute to heterochromatin establishment. Phylogenetic analysis resolves alpha satellite monomers within 450 kb region into four subfamilies, while monomers within DXZ1 array form distinct, fifths alpha satellite subfamily. Adjacent to 450 kb region is euchromatic DNA

an average of 1–2% divergence on the same or different X chromosome. Mapping of deletion chromosomes has delimited the functional centromere of the X chromosome to the higher order alpha satellite array in the DXZ1 region. The other class is composed of ~450 kb region located between DXZ1 and expressed sequences on the short arm of chromosome X, also highly enriched in alpha satellite. The 450 kb junction region is characterized by tandemly repeated monomeric repeat structure and the monomers exhibit higher mutual divergence relative to higher order repeats within DXZ1 region.

Based on the presence of interspersed LINE elements within arrays of alpha satellite DNA as well as on the phylogenetic analysis of primate species, particular alpha satellite subdomains can be defined and their age can be estimated. According to such analyses, human X chromosome monomeric alpha satellite arrays are divided into four age groups: 35–65 million years (Myr), 25–35 Myr, 15–25 Myr, 7–15 Myr, while the DXZ1 region which is based on higher order repeats is the most recent one with an approximate age between 2 and 7 Myr (Schueler and Sullivan 2006). Monomeric alpha satellite DNA predates higher order arrays of alpha satellite and may represent direct descendants of the ancestral primate centromere sequence. Comparison with centromeric alpha satellite DNA sequences in other primate species revealed that alpha satellite DNA has evolved through proximal expansion events occurring within the central active region of the centromere (Fig. 3.3; Schueler et al. 2005).

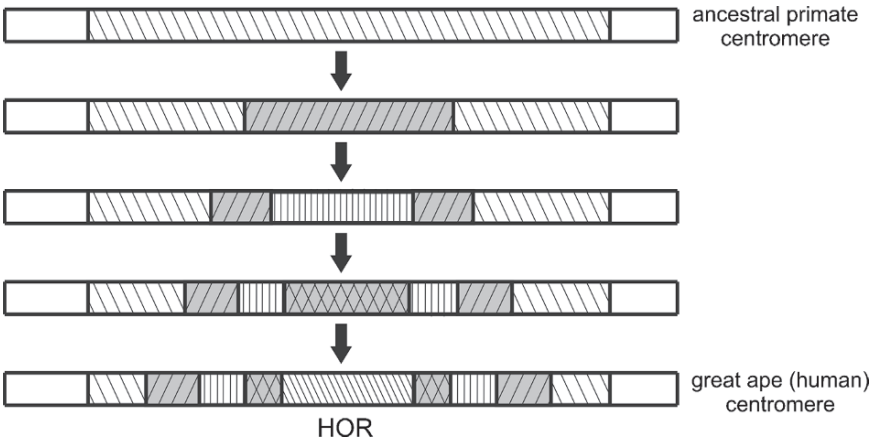


Fig. 3.3 Model of evolution of primate centromeric region from the ancestral primate to humans. The series of amplification events are responsible for the spreading of "new" alpha satellite subfamilies and replacement of "old" ones, which however remain preserved in genome in lower number of copies (differently dashed rectangles). In each round of amplification, the "old" centromere is split and moved distally onto each arm while the newly added sequence confers centromere function. The "old" subfamilies are based on tandemly repeated monomers, but the most recently amplified subfamily is based on tandemly repeated HOR. This subfamily comprises centromeric regions in humans and other great ape. The model is based on data on human X chromosome centromere structure (Schueler et al. 2005)

Each addition of new material splits the previous centromeric DNA and moves it distally onto each arm, while the newly added sequence confers centromere function. The alpha satellite region immediately proximal to the euchromatin chromosome arm is a remnant of the ancestral primate X centromere. A higher order satellite array located within the DXZ1 domain evolved as a replacement for the monomeric alpha satellite repeat. Highly homogenous arrays of higher-order alpha satellite represent a relatively recent addition to the primate genome, emerging near the orangutan/gorilla split. Based on the molecular analysis of the human X-chromosome centromere, it becomes evident that alpha satellite regions have evolved through a series of events, resulting in the addition and amplification of "new" subfamilies that have partially replaced the "old" ones (Fig. 3.3).

The kinetochore domain composed of higher order repeats comprises one half to two thirds of the alpha satellite DNA located at human centromeres. The remainder of alpha satellite arrays composed predominantly of diverged tandemly repeated monomers contributes to pericentromeric heterochromatin establishment, which is necessary for chromatid cohesion.

3.5.2 Model of Centromere Evolution Based on Satellite DNA Library

Rapid sequence evolution is characteristic of complex regional centromeres. Comparison of alpha satellite arrays from orthologous chromosomes of chimps and human revealed higher divergence of centromeric regions relative to the pericentromeric ones (Rudd et al. 2006). To explain rapid evolution of centromeric DNA, a «centromere drive hypothesis» has been introduced (Malik and Henikoff 2002; see Chap. 2 in this book). According to it, rapid evolution of centromeric DNA is caused by positive selection that imposes a bias in favour of retaining mutations in centromere region. The positive selection is proposed to be due to the advantage conferred to mutated centromere during female meiosis. Such centromere has a higher affinity for centromeric chromatin proteins and is the most successful at being incorporated into the functional germ cells (i.e., the oocyte). Other centromeres are then forced to adopt the same sequence and protein variants to segregate efficiently. According to the "centromere drive hypothesis," evolution of the centromere proceeds through «de novo» adoption of «new», previously noncentromeric sequences that are repeatedly introduced into the genome (Dawe and Henikoff 2006).

On the other hand, based on the library hypothesis, it can be proposed that centromere is formed from already adapted sequences with certain structural characteristics that enable them to confer a centromeric role or to perform some other function such as regulation of gene expression (Ugarković 2005; 2008b; Fig. 3.1). Such sequences after exaptation, that is, after becoming functional, can reside within the genome for long evolutionary periods and create a satellite DNA library. The content of the library is constantly evolving, and new sequences can be generated and added into the library such as the case of alpha satellite complex HORs,

which appear later in the evolution of primate lineage (Alexandrov et al. 2001). On the other hand, some «old» centromeric satellite repeats can be lost in particular lineages as shown for centromeric satellites in species of grass (Lee et al. 2005). Removal of centromeric satellites from the library is probably a stochastic process mediated by mechanisms of unequal crossing over and illegitimate recombination (Stephan 1986; Ma and Jackson 2006).

Centromeric and pericentromeric satellite sequences from the library can undergo recurrent repeat copy number expansion and contraction in divergent lineages (Fig. 3.4). Such changes in copy number seem to be random and do not correlate

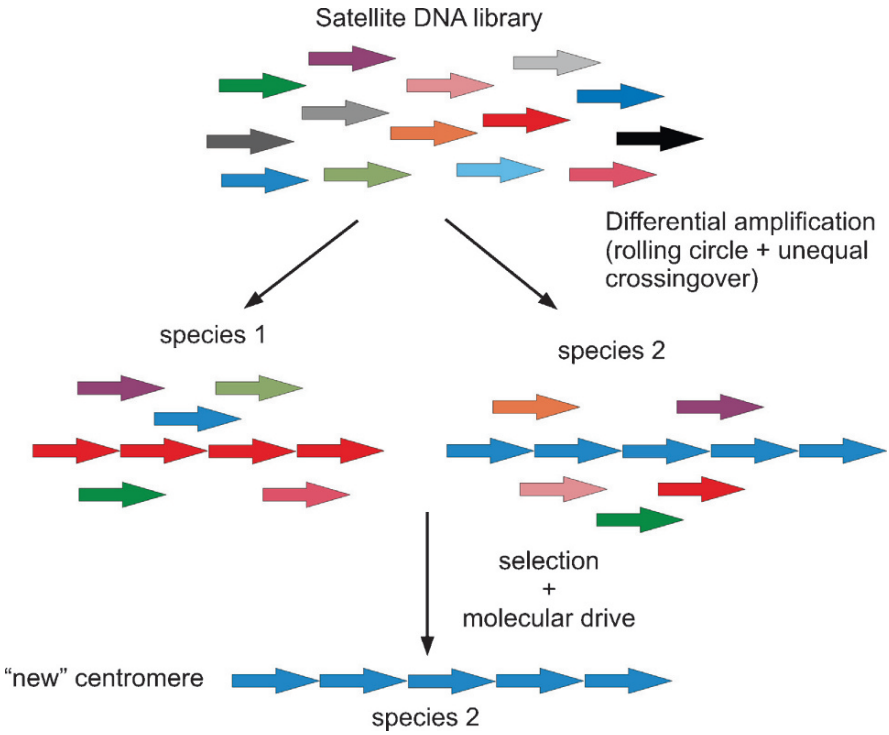


Fig. 3.4 Model of satellite DNA evolution and centromere formation based on satellite DNA library. Satellite DNAs possessing certain structural features which enable them to become functional are retained in the genome in the form of satellite DNA library. Satellite DNA could have dual function in the genome: either it can be extended into long array and together with its transcripts participates in centromere/kinetochore establishment, or satellite transcripts could act as regulators of gene expression, probably through RNAi mechanism. A stochastic process of differential amplification of satellite DNAs from the library in two related species induced by unequal crossingover, duplicative transposition or extrachromosomal rolling circle replication can lead to the formation of long, uninterrupted arrays. An expanded arrays can replace the previous centromere if it has some selective advantage relative to the «old» centromere, e.g., transmission advantage at meiosis due to some structural characteristic or just due to the higher homogeneity of newly amplified array relative to the «old» one. Such "new" centromere can then be spread through the population by processes of natural selection and molecular drive

with phylogeny of the species as shown for the insect genus *Pimelia*, the marsupial genus *Macropus*, and the grass species (Pons et al. 2004; Lee et al. 2005; Bulazel et al. 2007). The same satellite sequences can undergo convergent expansion on all chromosomes in different lineages. Although the evolution of centromeric satellite DNA composition does not follow species phylogeny, it parallels chromosome evolution in some karyotypically divergent lineages (Slamovits et al. 2001; Bulazel et al. 2007; see Chap. 4 in this book). The rate of turnover of centromere differs among species ranging from abrupt-saltatory amplification and replacement of "old" centromere in relatively short periods of time, through gradual changes, while in some instances no apparent change occurs for long evolutionary time (Pons et al. 2004). Amplification of a satellite sequences could occur due to unequal crossingover or duplicative transposition (Smith 1976; Ma and Jackson 2006), while the spreading and fixation in population can be influenced by stochastic process of molecular drive (Dover 1986) and by natural selection. The discovery of human extrachromosomal elements originating from satellite DNA arrays in cultured human cells indicates the possible existence of other amplification mechanisms based on extrachromosomal rolling-circle replication (Assum et al. 1993). Satellite DNA-derived extrachromosomal circular DNA is common in plant genomes and is considered as an intermediate in process driving satellite expansion and evolution (Navratilova et al. 2008). It has been proposed that satellite sequences excised from their chromosomal loci via intrastrand homologous recombination could be amplified in this way, followed by reintegration of tandem arrays into the genome (Feliciello et al. 2006). Mechanistic processes inherent to chromosome fusion and translocation have also been supposed to be responsible for contraction and expansion of centromeric satellite DNA arrays (Bulazel et al. 2007).

A newly expanded satellite array can replace the previous centromere and prevail in the population if it has some selective advantage relative to the «old» centromere, for example, transmission advantage at meiosis due to some sequence or structural characteristic of newly amplified satellite DNA or just due to the higher homogeneity of newly amplified array relative to the «old» one (Fig. 3.4). Based on the structure of the human X chromosome centromere, it can be proposed that high homogeneity and integrity of newly expanded satellite arrays might represent an additional requirement imposed on the centromere. In addition, it seems that a newly expanded array has to be of certain length to become a preferred substrate for centromere formation. This could be related to the number of microtubule binding sites per chromosome necessary to ensure the proper chromosome segregation.

The repetitiveness of satellite DNA has been proposed to be important for orderly packing of nucleosomes (Vogt 1990), and nucleosome crystallization on reverse repeats of alpha satellite DNA support this assumption (Harp et al. 1996; Luger et al. 1997). There is strong indication that a specific set of periodic DNA motifs encoded in tandemly repeated satellite DNA provides signals for specific chromatin organization in the form of distinctive nucleosome arrays characteristic for centromere (Takasuka et al. 2008). It is known that centromeric nucleosomes are organized as a heterotypic tetramer composed of one molecule each of CenH3, H2A, H2B, and H4, different from the octamer found in bulk nucleosomes

(Dalal et al. 2007). It is suggested that such nucleosome tetramers distributed orderly on homogenous and uninterrupted satellite arrays represent an accessible surface for kinetochore assembly. Therefore, extension of satellite repeat from the library by stochastic recombinational processes and/or extrachromosomal rolling circle replication might create uninterrupted homogenous array, which could be a favoured substrate for centromere chromatin establishment and microtubule binding relative to the "old" nonhomogenous array interspersed with different transposable elements. Such centromere array exhibiting a slight advantage relative to the "old" one could then be fixed in a population (Fig. 3.4).

3.6 RNA in Centromere Establishment

3.6.1 RNAs as Epigenetic Regulator of Heterochromatin Establishment

Transcripts of centromeric satellite DNAs have been reported in several organisms, including vertebrates, invertebrates, and plants. Transcripts are usually heterogeneous in size and are in some cases strand-specific, while in others transcription proceeds from both DNA strands. Most transcripts are present as polyadenylated RNA in the cytoplasm but some are found exclusively in the nucleus (reviewed in Ugarković 2005). Recently, it has been shown that transcripts derived from tandemly repeated centromeric DNA of the fission yeast *S. pombe* exist in the form of small 20–25 bp long RNAs that are involved in chromatin modifications and establishment of heterochromatin (Volpe et al. 2002). The chromatin silencing mechanism is initiated by long double-stranded RNA (dsRNA) that arises from bidirectional transcription of repeated centromeric DNA and is further processed by the RNAse III-like ribonuclease Dicer into small interfering RNAs (siRNAs). siRNAs are then loaded into the RNA-induced transcriptional silencing complex (RITS) through their association with the Argonaute protein. RITS also interacts with the RNA-directed RNA polymerase complex (RDRC), which is required for the production of secondary dsRNA and amplification of the silencing signal (Verdel et al. 2004). Both RITS and RDRC associate with the nascent noncoding centromeric RNA transcript, and binding to RITS is probably achieved through the base-pairing of siRNA molecules with nascent RNA and by direct contact with the RNA pol II elongation complex. In addition to siRNAs, the association of RITS with chromatin also requires a histone methyltransferase. Histone H3 methylation at lysine 9 is essential for the recruitment of heterochromatin protein 1 (HP1). This represents an initial step in the formation of heterochromatin. HP1 has several functions at centromere such as silencing gene expression and recombination, promotion of kinetochore assembly, and prevention of erroneous microtubule attachment to the kinetochores (Yamagishi et al. 2008).

Mutations in components of the RNAi pathway lead to the loss of pericentromeric heterochromatin in fission yeast, resulting in mis-segregation of chromosomes

(Allshire et al. 1995; Volpe et al. 2002; Fig. 3.5). *S. pombe* cells deficient in pericentromeric heterochromatin are unable to recruit the chromosome cohesin to centromeres and fail to maintain centromere cohesion (Bernard et al. 2001). It was recently revealed that heterochromatic proteins and RNAi machinery promote CENP-A deposition and kinetochore assembly over the central domain of the fission yeast centromere (Folco et al. 2008). However, absence of these factors does not affect CENP-A deposition on endogenous centromeres or on minichromosome centromeres, which have incorporated CENP-A in previous generation. In general, pericentromeric heterochromatin appears to be an absolute requirement for the establishment of centromere in fission yeast together with central DNA region, which binds CENP-A (*cnt* region) as well as *otr* region which contains *dg-dh* repeats (Folco et al. 2008). In addition to fission yeast, pericentromeric heterochromatin seems to be required for the accurate segregation of chromosomes during mitosis in many eukaryotes, including *Drosophila* and mammals (Kellum and Alberts 1995; Peters et al. 2001).

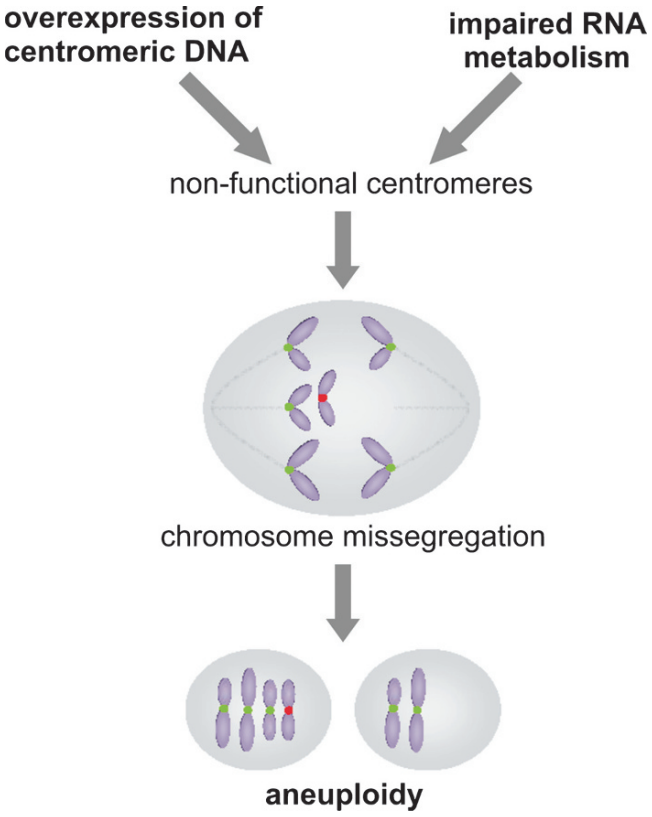


Fig. 3.5 Link between centromeric RNA and aneuploidy. Aberrant expression of centromeric satellite DNA affects centromere/kinetochore function and causes abnormality in chromosome segregation. Defects in RNA metabolism could affect heterochromatin maintenance and fidelity in mitosis

RNA interference (RNAi) machinery has been shown to be evolutionary conserved and is proposed to be responsible for pericentromeric heterochromatin formation in different animal species. In addition to *S. pombe*, siRNAs cognate to satellite DNAs are involved in the epigenetic process of chromatin modification in *Arabidopsis* and *C. elegans* (Bernstein and Allis 2005; Grewal and Elgin 2007). In *D. melanogaster* RNAi seems to be involved in the establishment of heterochromatin in early embryo. Once set, heterochromatin can be maintained in the absence of RNAi in somatic tissues (Huisinga and Elgin 2008). In mammals, however, siRNAs seem not to elicit chromatin modification, although an unidentified RNA component appears to be required for maintaining pericentric heterochromatin (Maison et al. 2002; Wang et al. 2006). In mouse pericentromeric heterochromatin, γ satellite DNA as its major constituent is transcribed as small, approximately 200-nt-long RNA during mitosis, while during G1 and S phase, transcription occurs in the form of long, heterogeneous RNAs (Lu and Gilbert 2007). The transcription is cell-cycle regulated with the highest rate in early S phase and in mitosis, similar to regulation in fission yeast where the peak of transcription occurs at S phase (Chen et al. 2008). Besides being cell-cycle regulated, transcription of mouse pericentromeric heterochromatin is also linked to cellular proliferation.

3.6.2 RNAs as Structural Component of Centromere

Recently it has been shown that long, single-stranded alpha satellite DNA transcripts encompassing a few satellite monomers are functional components of the human kinetochore (Wong et al. 2007; Fig. 3.1). Centromere alpha satellite RNA is required for the assembly of CENPC1, INCENP (inner centromere protein), and survivin (an INCENP-interacting protein) at the metaphase centromere. It also directly facilitates the accumulation and assembly of centromere-specific nucleoprotein components at the interphase nucleolus. The nucleolus sequesters centromeric components such as alpha satellite RNA and centromere proteins for timely delivery to the chromosomes for kinetochore assembly at mitosis. CENP-C has been shown to be an RNA-associating protein that binds alpha satellite RNA, as revealed by in vitro binding assay. The same protein also binds alpha satellite DNA in vivo and obviously has dual RNA- and DNA-binding function (Politi et al. 2002). In mammals, CENP-C evolving rapidly and different from CENP-A (vertebrate CenH3) shows evidence of positive selection (Talbert et al. 2004; see Chap. 2 in this book). It is possible that a pool of CENP-C has a centromere DNA-binding role that persists throughout the cell cycle. The other pool of CENP-C is involved in relocation of alpha satellite RNA and centromere proteins from the nucleolus onto the mitotic centromere.

CENP-B and CENP-C recognize the same subfamilies of alpha satellite DNA, but it is not clear whether CENP-C preferentially recognizes a specific sequence within satellite DNA or RNA. In vitro experiments indicate that CENP-C does not bind a specific DNA sequence, similar to CENP-A which also seems to be a sequence nonspecific binding protein (Politi et al. 2002). However, the existence of

binding sites for different proteins in alpha satellite DNA could explain the nonrandom distribution of mutations within a sequence and can give strong support for the influence of selection on the evolution of this satellite DNA sequence.

Numerous examples illustrate the involvement and possible importance of longer RNAs for the formation of centromeric chromatin and for centromere function. RNA encoded by centromeric satellite DNA and retrotransposons, ranging in size between 40 and 200 nt, has been shown to be an integral component of the kinetochore in maize, tightly bound to centromeric histone H3 (Topp et al. 2004). Murine minor satellite DNA associated with the centromeric region is transcribed from both strands, and transcripts are processed into 120 nt RNA, which localizes to the centromere (Bouzinba-Segard et al. 2006). The overexpression of satellite transcripts is impaired by mislocalization of centromere-associated proteins essential for the formation of centromeric heterochromatin. In addition, forced accumulation of transcripts leads to defects in chromosome segregation and impaired centromere function, resulting in aneuploidy (Fig. 3.5). The absence of siRNAs homologous to murine minor satellite indicates that the longer noncoding RNA plays a role in heterochromatin formation and centromere establishment in the murine system. Long, stable transcripts of centromeric satellite DNAs are also the characteristics of some beetle species (Pezer and Ugarković 2008a; 2009). Functional studies reveal that in this animal system an increase in the amount of centromeric satellite DNA transcripts coincides with the irregular chromosome segregation and often leads to aneuploidy. Since functional promoters for RNA polymerase II are detected within satellite DNAs from coleopteran genera *Tribolium* and *Palorus*, it is proposed that constitutive expression of centromeric satellites is necessary for proper centromere establishment (Pezer and Ugarković 2008b).

Mitotic and chromosome segregation defects have been reported for fission yeast mutants defective in RNA metabolism (Win et al. 2006). RNase activity of Dis3, a core component of the exosome that is required for the processing of different RNAs, is shown to be required for heterochromatin silencing within the centromere as well as for proper kinetochore formation and establishment of kinetochore–microtubule interactions (Murakami et al. 2007; Buhler et al. 2007). Thus, RNAi-independent degradation of centromeric transcripts also contributes to heterochromatin formation and proper centromere function.

All these examples demonstrate the importance of cellular RNA metabolism for proper chromosome segregation during mitosis (Fig. 3.5). In addition to the relatively well understood RNAi mechanism that moderates heterochromatin establishment in different eukaryotic systems, other mechanisms involving longer RNAs also operate in centromeric chromatin assembly and kinetochore formation. Although these mechanisms are poorly understood, it can be proposed that centromere-encoded longer RNAs could serve as a scaffold for chromatin-remodeling complexes at centromere as well as structural component of kinetochore (Fig. 3.1). It can be proposed that specific secondary and tertiary structures of centromeric RNAs are important for assembly of such complexes.

Based on studies in mammalian and insect systems, it appears that aberrant transcription of noncoding centromeric satellite DNA affects heterochromatin

maintenance and fidelity of mitosis (Pezer and Ugarković 2008b; Frescas et al. 2008). This indicates that centromeric RNA is an important functional component of the centromere/kinetochore complex, probably tightly bound to proteins, and subtle changes in centromeric RNA/kinetochore protein ratio affect chromosome stability and segregation (Fig. 3.5). Stoichiometric expression of all kinetochore components including proteins and noncoding centromeric RNA seems to be important for normal kinetochore assembly and function.

Overexpression of noncoding satellite DNAs is characteristic of some tumours. Analysis of transcription of human satellite 2 and α-satellite, which are located in pericentromeric and centromeric heterochromatin, respectively, revealed an elevated level of their expression in ovarian epithelial carcinomas and Wilms tumours, relative to the control (Alexiadis et al. 2007). It can be hypothesized that increased accumulation of noncoding RNA deriving from the two satellite DNAs interferes with heterochromatin formation and kinetochore establishment, affecting in this way mitotic segregation.

3.7 Conclusion

It can be proposed that the occurrence of new centromere results from a stochastic process affecting repetitive DNA, which is induced by homologous recombination followed probably by extrachromosomal rolling circle replication. As a result of such process, amplification of different satellite sequences already present within a genome occurs. However, only those satellites that have inherent centromere-competence in the form of some structural requirements necessary for centromere function are after amplification fixed in a population as a new centromere.

Presence of some conserved structural motifs within satellite DNAs such as periodically distributed AT tracts or protein binding sites indicates that despite centromere sequence flexibility, there are structural determinants that are prerequisite for centromere function. In addition, detection of transcripts from centromeric DNA that represent structural component of centromere indicates possible importance of structural elements at the level of RNA secondary or tertiary structures.

Acknowledgements This work was supported by grant 00982604 from the Croatian Ministry of Science and EU FP6 Marie Curie Transfer of Knowledge Grant MTKD-CT-2006-042248. The author is grateful to Josip Brajković for the help with figures.

References

Alexandrov I, Kazakov A, Tumeneva I, Shepelev V, Yurov Y (2001) Alpha-satellite DNA of primates: old and new families. Chromosoma 110:253–266

Alexiadis V, Ballestas ME, Sanchez C, Winokur S, Vedanarayanan V, Warren M, Ehrilch M (2007) RNAPol-ChIP analysis of transcription from FSHD-linked tandem repeats and satellite DNA. Biochim Biophys Acta 1796:29–40

Allshire RC, Nimmo ER, Ekwall K, Javerzat JP, Cranston G (1995) Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. Genes Dev 9:218–233
Amor DJ, Choo KH (2002) Neocentromeres: role in human disease, evolution and centromere studies. Am J Hum Genet 71:695–714
Assum G, Fink T, Steinbeisser T, Fisel KJ (1993) Analysis of human extrachromosomal DNA elements originating from different beta-satellite subfamilies. Hum Genet 91:489–495
Baker RE, Rogers K (2005) Genetic and genomic analysis of the AT-rich centromere DNA element II of *Saccharomyces cerevisiae*. Genetics 171:1463–1475
Basu J, Stromberg G, Compitello G, Willard HF, Van Bokkelen G (2005) Rapid creation of BAC-based human artificial chromosome vectors by transposition with synthetic alpha-satellite arrays. Nucleic Acids Res 33:587–596
Bensasson D, Zarowiecki M, Burt A, Koufopanou V (2008) Rapid evolution of yeast centromeres in the abscence of drive. Genetics 178:2161–2167
Bernard P, Maure JF, Partridge JF, Genier S, Javerzat JP, Allshire RC (2001) Requirement of heterochromatin for cohesion at centromeres. Science 21:2539–2542
Bernstein E, Allis CD (2005) RNA meets heterochromatin. Genes Dev 19:1635–1655
Black BE, Bassett EA (2008) The histone variant CENP-A and centromere specification. Curr Opin Cell Biol 20:91–100
Blower MD, Nachury M, Heald R, Weis K (2005) A Rac-1 containing ribonucleoprotein is required for mitotic spindle assemby. Cell 121:223–234
Bouzinba-Segard H, Guais A, Francastel C (2006) Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proc Natl Acad Sci USA 103:8709–8714
Bruvo-Mađarić B, Plohl M, Ugarkovi Đ (2007) Wide distribution of related satellite DNA families within the genus *Pimelia* (Tenebrionidae). Genetica 130:35–42
Bulazel KV, Ferreri GC, Eldridge MD, O' Neill RJ (2007) Species-specific shifts in centromere sequence composition are coincident with breakpoint reuse in karyotypically divergent lineages. Genome Biol 8:R170
Bühler M, Haas W, Gygi SP, Moazed D (2007) RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. Cell 129:707–721
Cesari M, Luchetti A, Passamonti M, Scali V, Mantovani B (2003) Polymerase chain reaction amplification of the Bag320 satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). Gene 312:289–295
Charlesworth B, Langley CH, Stephan W (1986) The evolution of restricted recombination and the accumulation of repeated DNA sequences. Genetics 112:947–962
Chen ES, Zhang K, Nicolas E, Cam HP, Zofall M, Grewal SI (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature 451:734–737
Clarke L, Carbon J (1983) Genomic substitutions of centromeres in *Saccharomyces cerevisiae*. Nature 305:23–28
Coats SR, Zhang Y, Epstein LM (1994) Transcription of satellite 2 DNA from the newt is driven by a snRNA type of promoter. Nucleic Acids Res 22:4697–4704
Cooper JL, Henikoff S (2004) Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol 21:1712–1718
Dalal Y, Furuyama T, Vermaak D, Henikoff S (2007) Structure, dynamics, and evolution of centromeric nucleosomes. Proc Natl Acad Sci USA. 104:15974–15981
Dawe RK, Henikoff S (2006) Centromeres put epigenetics in the driver's seat. Trends Biochem Sci 31:662–669
Dobie KW, Hari KL, Maggert KA, Karpen GH (1999) Centromere proteins and chromosome inheritance: a complex affair. Curr Opin Genes Dev 9:206–217
Dover GA (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 2:159–165
Durajlija žini S, Ugarkovi Đ, Cornudella L, Plohl M (2000) A novel interspersed type of organization of satellite DNAs in *Tribolium madens* heterochromatin. Chromosome Res 8:201–212
Feliciello I, Picariello O, Chinali G (2005) The first characterisation of the overall variability of repetitive units in a species reveals unexpected features of satelite DNA. Gene 349:153–164

Feliciello I, Picariello O, Chinali G (2006) Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from *Rana dalmatina*: molecular evidence of a new mechanism of DNA repair acting on satellite DNA. Gene 383:81–92

Ferbeyre G, Smith JM, Cedergren R (1998) Schistosome satellite DNA encodes active hammerhead-ribozymes. Mol Cell Biol 18:3880–3888

Fitzgerald DJ, Dryden GL, Bronson EC, Williams JS, Anderson JN (1994) Conserved pattern of bending in satellite and nucleosome positioning DNA. J Biol Chem 269:21303–21314

Folco HD, Pidoux AL, Urano T, Allshire RC (2008) Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. Science 319:94–97

Frescas D, Guardavaccaro D, Kuchay SM, Kato H, Poleshko A, Basrur V, Elenitoba-Johnson KS, Katz RA, Pagano M (2008) KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. Cell Cycle 7(29):3539–3547

Fry K, Salser W (1977) Nucleotide sequences of HS-α satellite DNA from kangaroo rat *Dipodomys ordii* and characterisation of similar sequences in other rodents. Cell 12:1069–1084

Grewal SI, Elgin SC (2007) Transcription and RNA interference in the formation of heterochromatin. Nature 447:399–406

Grimes BR, Rhoades AA, Willard HF (2002) Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. Mol Ther 5:798–805

Hall SE, Kettler G, Preuss D (2003) Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. Genome Res 13:195–205

Harp JM, Uberbacher EC, Roberson AE, Palmer EL, Gewiess A, Bunick GJ (1996) X-ray diffraction analysis of crystals containing twofold symmetric nucleosome core particles. Acta Crystallogr D 52:283–288

Hegemann JH, Fleig UN (1993) The centromere of budding yeast. Bioessays 15:451–460

Henikoff S, Dalal Y (2005) Centromeric heterochromatin: what makes it unique. Curr Opin Genet Dev 15:177–184

Heslop-Harrison JS, Murata M, Ogura Y, Schwarzacher T, Motoyoshi F (1999) Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes. Plant Cell 11:31–42

Huisinga KL, Elgin SCR (2009) Small RNA-directed heterochromatin formation in the context of development: What flies might learn from fission yeast. Biochim Biophys Acta 1789:3–16

Jaco I, Canela A, Vera E, Blasco MA (2008) Centromere mitotic recombination in mammalian cells. J Cell Biol 181:885–92

Jin W, Melo JR, Nagaki K, Talbert PB, Henikoff S, Dawe RK, Jiang J (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. Plant Cell 16:571–581

Jonstrup AT, Thomsen T, Wang Y, Knudsen BR, Koch J, Andersen AH (2008) Hairpin structures formed by alpha satellite DNA of human centromeres are cleaved by human topoisomerase II α. Nucleic Acids Res 36:6165–6175

Kalitsis P (2008) Centromeres. In: Encyclopedia of life sciences (ELS). Wiley, Chichester

Kawabe A, Charlesworth D (2007) Patterns of DNA variation among three centromere satellite families in *Arabidopsis halleri* and *A. lyrata*. J Mol Evol 64:237–247

Kellum R, Alberts BM (1995) Heterochromatin protein 1 is required for correct chromosome segregation in *Drosophila embryos*. J Cell Sci 108:1419–1431

King K, Jobst J, Hemleben V (1995) Differential homogenisation and amplification of two satellite DNAs in the genus *Cucurbita* (Cucurbitaceae). J Mol Evol 4:996–1005

Kipling D, Warburton PE (1997) Centromeres, CENP-B and *Tigger* too. Trends Genet 13:141–145

Kumekawa N, Hosouchi T, Tsuruoka H, Kotani H (2001) The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 4. DNA Res 8:285–290

Lee C, Wevrick R, Fisher RB, Ferguson-Smith MA, Lin CC (1997) Human centromeric DNAs. Hum Genet 100:291–304

Lee HR, Zhang W, Langdon T, Jin W, Yan H, Cheng Z, Jiang J (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. Proc Natl Acad Sci USA 102:11793–117998

Li YX, Kirby ML (2003) Coordinated and conserved expression of alphoid repeat and alphoid repeat-tagged coding sequences. Dev Dynamics 228:72–81

Lin CC, Li YC (2006) Chromosomal distribution and organization of three cervid satellite DNAs in Chinese water deer (Hydropotes inermis). Cytogenet Genome Res 114:147–154
Lu J, Gilbert DM (2007) Proliferation-dependent and cell cycle-regulated transcription of mouse pericentromeric heterochromatin. J Cell Biol 179:411–421
Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8-angstrom resolution. Nature 389:251–260
Ma J, Jackson SA (2006) Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. Genome Res 16:251–259
Maddox PS, Oegema K, Desai A, Cheesman IM (2004) "Holo"er than thou: chromosome segregation and kinetochore function in *C. elegans*. Chromosome Res 12:641–653
Maison C, Bailly D, Peters AH, Quivy JP, Roche D, Taddei A, Lachner M, Jenuwein T, Almouzni G (2002) Higher-order structure in pericentromeric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat Genet 30:329–334
Malik HS, Henikoff S (2002) Conflict begets complexity: the evolution of centromeres. Curr Opin Genet Dev 12:711–718
Marshall OJ, Chuch AC, Wong LH, Choo KH (2008) Neocentromeres. new insights into centromere structure, disease, development, and karyotype evolution. Am J Hum Genet 82:261–282
Martinez-Balbas A, Rodriguez-Campos A, Gracia-Ramirez M, Sainz J, Carrera P, Aymami J, Azorin F (1990) Satellite DNAs contain sequences that induce curvature. Biochemistry 29:2342–2348
Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromere satellite. J Cell Biol 109:1963–1973
Masumoto H, Nakano M, Ohzeki J (2004) The role of CENP-B and alpha-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres. Chromosome Res 12:543–556
Meraldi P, McAinsh AD, Rheinbay E, Sorger PK (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. Genome Biol 7:R23
Meštrović N, Plohl M, Mravinac B, Ugarković, Đ (1998). Evolution of satellite DNAs from the genus *Palorus*- experimental evidence for the "library" hypothesis. Mol Biol Evol 15:1062–1068
Meštrović N, Castagnone-Sereno P, Plohl M (2006) Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. Mol Biol Evol 23:2316–2325
Metz A, Soret J, Vourc'h C, Tazi J, Jolly C (2004) A key role for stress-induced satellite III transcripts in the relocalization of splicing factors into nuclear stress granules. J Cell Sci 117:4551–4558
Mravinac B, Plohl M, Meštrović N, Ugarković Đ (2002) Sequence of PRAT satellite DNA "frozen" in some coleopteran species. J Mol Evol 54:774–783
Mravinac B, Plohl M, Ugarković Đ (2004) Conserved patterns in the evolution of *Tribolium* satellite DNAs. Gene 332:169–177
Mravinac B, Plohl M, Ugarković Đ (2005) Preservation and high sequence conservation of satellite DNAs suggest functional constraints. J Mol Evol 61:542–550
Murakami H, Goto DB, Toda T, Chen ES, Grewal SI, Martienssen RA, Yanagida M (2007) Ribonuclease Activity of Dis3 is required for mitotic progression and provides a possible link between heterochromatin and kinetochore function. PLoS ONE 3:e317
Nasmyth K (2002) Segregating sister genomes: the molecular biology of chromosome separation. Science 288:559–565
Navratilova A, Koblizkova A, Macas J (2008) Survey of extrachromosomal circular DNA derived from plant satellite repeats. BMC Plant Biol 8:90
Nijman IJ, Lenstra JA (2001) Mutation and recombination in cattle satellite DNA: a feedback model for the evolution of satellite DNA repeats. J Mol Evol 52:361–371
Ohzeki J, Nakano M, Okada T, Masumoto H (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J Cell Biol 159:765–775
Okada T, Ohzeki J, Nakano M, Yoda K, Brinkley WR, Larionov V, Masumoto H (2007) CENP-B controls centromere formation depending on the chromatin context. Cell 131:187–1300

Pezer Ž, Ugarković Đ (2008a) RNA Pol II promotes transcription of centromeric satellite DNA in Beetles. PLoS ONE 3:e1594
Pezer Ž, Ugarković Đ (2008b) Role of non-coding RNA and heterochromatin in aneuploidy and cancer. Semin Cancer Biol 18:123–130
Pezer Ž, Ugarković Đ (2009) Transcription of pericentromeric heterochromatin in beetles – satellite DNAs as active regulatory elements. Cytogenet Genome Res (in press)
Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, Jenuwein T (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell 107:323–337
Politi V, Perini G, Trazzi S, Pliss A, Raska I, Earnshaw WC, Della Valle G (2002) CENP-C binds the alpha-satellite DNA in vivo at specific centromere domains. J Cell Sci 11:2317–2327
Pons J, Bruvo B, Petitpierre E, Plohl M, Ugarković D, Juan C (2004) Complex structural feature of satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): random differential amplification from a common "satellite DNA library". Heredity 92:418–427
Renault S, Roulex-Bonnin F, Periquet G, Bigot Y (1999) Satellite DNA transcription in *Diadromus pulchellus* (Hymenoptera). Insect Biochem Mol Biol 29:103–111
Romanova LY, Deriagin GV, Mashkova TG, Tumeneva IG, Mushegian AR, Kisselev LL, Alexandrov IA (1996) Evidence for selection in evolution of alpha satellite DNA: the central role of CENP-B/pJα binding region. J Mol Biol 261:334–340
Rudd MA, Wray GA, Willard HF (2006) The evolutionary dynamics of α-satellite. Genome Res 16:88–96
Sanyal K, Baum M, Carbon J (2004) Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique. Proc Natl Acad Sci USA 101:1134–11379
Schueler MG, Sullivan B (2006) Structural and functional dynamics of human centromeric heterochromatin. Annu Rev Genomics Hum Genet 7:301–313
Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. Science 294:109–115
Schueler MG, Dunn JM, Bird CP, Ross MT, Viggiano L; NISC Comparative Sequencing Program, Rocchi M, Willard HF, Green ED (2005) Progressive proximal expansion of the primate X chromosome centromere. Proc Natl Acad Sci USA 102:10563–10568
Slamovits CH, Cook JA, Lessa EP, Rossi MS (2001) Recurrent amplifications and deletions of satellite DNA accompanied chromosomal diversification in South American tuco-tucos (genus *Ctenomys*, Rodentia: Octodontidae): a phylogenetic approach. Mol Biol Evol 18:1708–1719
Smith PG (1976) Evolution of repeated sequences by unequal crossover. Science 191:528–535
Stephan W (1986) Recombination and the evolution of satellite DNA. Genet Res 47:167–174
Stephan W (2007) Evolution of genome organization. In: Encyclopedia of Life Sciences (ELS). Wiley, Chichester
Sun X, Le HD, Janice M, Wahlstrom JM, Karpen GH (2003) Sequence analysis of a functional *Drosophila* centromere. Genome Res 13:182–194
Tal M, Shimron F, Yagil G (1994) Unwound regions in yeast centromere IV DNA. J Mol Biol 243:179–189
Talbert PB, Bryson TD, Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. J Biol 3:18
Takasuka TE, Cioffi A, Stein A (2008) Sequence information encoded in DNA that may influence long-range chromatin structure correlates with human chromosome functions. PLoS ONE 3:e2643
Topp CN, Zhong CX, Dawe RK (2004) Centromere-encoded RNAs are integral components of the maize kinetochore. Proc Natl Acad Sci USA 101:15986–15991
Ugarković Đ (2005) Functional elements residing within satellite DNAs. EMBO Rep 6:1035–1039
Ugarković Đ (2008a) Evolution of Alpha satellite DNA. In: Encyclopedia of Life Sciences (ELS). Wiley, Chichester
Ugarković Đ (2008b) Satellite DNA libraries and centromere evolution. Open Evol J 2:1–6

Ugarković Đ, Plohl M (2002) Variation in satellite DNA profiles – causes and effects. EMBO J 21:5955–5959

Ugarković Đ, Podnar M, Plohl M (1996a) Satellite DNA of the red flour beetle *Tribolium castaneum*-comparative study of satellites from the genus *Tribolium*. Mol Biol Evol 13:1059–1066

Ugarković Đ, Durajlija S, Plohl M (1996b) Evolution of *Tribolium madens* (Insecta, Coleoptera) satellite DNA through DNA inversion and insertion. J Mol Evol 42:350–358

Ugarković ĐL, Plohl M, Lucijanić-Justić V, Borštnik B (1992) Detection of satellite DNA in *Palorus atzeburgii*: Analysis of curvature profiles and comparison with *Tenebrio molitor* satellite DNA. Biochimie 74:1075–1082

Vagnarelli P, Ribeiro SA, Earnshaw WC (2008) Centromeres: old tales and new tools. FEBS Lett 582:1950–1959

Verdel A, Jia S, Gerber S, Suglyama T, Gygi S, Grewal SI, Moazed D (2004) RNAi-mediated targeting of heterochromatin with the RITS complex. Science 303:672–676

Vershinin AV, Alkhimova EG, Heslop-Harrison JS (1996) Molecular diversification of tandemly organised sequences and heterochromatic chromosome regions in some *Triticeae* species. Chromosome Res 4:517–525

Vogt P (1990) Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". Hum Genet 84:301–336

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297:1833–1837

Wang F, Koyama N, Nishida H, Haraguchi T, Reith W, Tsukamoto T (2006) The assembly and maintenance of heterochromatin initiated by transgene repeats are independent of the RNA interference pathway in mammalian cells. Mol Cell Biol 26:4028–4040

Westermann S, Drubin DG, Barnes G (2007) Structures and functions of yeast kinetochore complexes. Ann Rev Biochem 76:563–592

Win TZ, Stevenson AL, Wang SW (2006) Fission yeast Cid12 has dual functions in chromosome segregation and checkpoint control. Mol Cell Biol 26:4435–4447

Wong LH, Brettingham-Moore KH, Chan L, Quach JM, Anderson MA, Northrop EL, Hannan R, Saffery R, Shaw ML, Williams E, Choo KHA (2007) Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res 17:1146–1160

Yamagishi Y, Sakuno T, Shimura M, Watanabe Y (2008) Heterochromatin links to centromeric protection by recruiting shugoshin. Nature 455:251–256

Zhang Y, Huang YC, Zhang L, Li Y, Lu TT, Lu YQ, Feng Q, Zhao Q, Cheng ZK, Xue YB, Wing RA, Han B (2004) Structural features of the rice chromosome 4 centromere. Nucleic Acids Res 32:2023–2030

Zhu L, Chou SH, Reid BR (1996) A single G-to-C change causes human centromere TGGAA repeats to fold back into hairpins. Proc Natl Acad Sci USA 93:12159–12164

Chapter 4
The Role of ncRNA in Centromeres: A Lesson from Marsupials

Rachel J. O'Neill and Dawn M. Carone

Contents

Abstract Though centromeres have been thought to be comprised of repetitive, transcriptionally inactive DNA, new evidence suggests that eukaryotic centromeres produce a variety of transcripts and that RNA is essential for centromere competence. It has been proposed that centromere satellite transcripts play an essential role in centromere function through demarcation of the kinetochore-binding domain. However, the regional limits and regulation of transcription within the mammalian centromere are unknown. Analysis of transcriptional domains within the centromere in mammalian models is impeded by the unbridgeable expanse of satellite monomers throughout the pericentromere. The comparatively small size of the wallaby centromere and the evolutionary role of the centromere in marsupial speciation events position the wallaby centromere as a tractable and valuable mammalian centromere model. We highlight the current understanding of the wallaby centromere and the role of transcription in centromere function.

R.J. O'Neill (✉) and D.M. Carone
Center for Applied Genetics and Technology, Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA
e-mail: rachel.oneill@uconn.edu

Đ. Ugarković (ed.), *Centromere,* Progress in Molecular and Subcellular Biology 48,
DOI: 10.1007/978-3-642-00182-6_4, © Springer-Verlag Berlin Heidelbarg 2009

4.1 Centromere Structure

4.1.1 Genetic Components of the Centromere

The centromere, often seen as a primary constriction on the chromosome (excluding holocentric chromosomes where an entire chromatid can function as a centromere), is the site of sister chromatid separation facilitated by spindle fiber attachment. While the function of spindle attachment at the point of kinetochore formation is conserved across eukaryotes, the sequence and structure of domains adjacent to the centromere (the pericentromere) as well as the sequences found within the centromere proper (the core) are highly variable and remarkably divergent. This "centromere paradox" (Henikoff et al. 2001; see Sect. 4.2) has posed a challenge both to identifying the underlying features, organization, or structure *conserved* within centromeres of distantly related taxa and to defining the minimal requirements for proper centromere function.

A common feature of centromeres in higher eukaryotes is the presence of satellite DNA in both the core and the pericentric regions. While satellite DNA families can be species-specific (Singer 1982), their seemingly ubiquitous presence at or near centromeric domains suggests that they play a role in centromere function (Willard 1990; Eichler 1999; Henikoff et al. 2001). Fiber fluorescence in situ hybridization (FISH), the hybridization of specific probe sequences directly to mechanically stretched DNA fibers, is a powerful method to demarcate target sequence size and order along the length of a single DNA strand. Coupled with long-range mapping, these analyses have revealed that the blocks of centromere repeats range from 2 to 5 Mb in human and from 6 to 20 Mb in mouse (Choo 1997a). Similarly large centromere sizes have been estimated for several other higher eukaryotes (Choo 1997a, b; Li et al. 2000). However, studies of the DNA in neocentromeres (new centromeres formed in ectopic locations) in humans (du Sart et al. 1997; Sullivan and Willard 1998; Barry et al. 1999) and *Drosophila* (Williams et al. 1998) have shown that classical satellites are absent from these locations. Thus, satellite DNA may be sufficient for centromere function, but it is not required (Willard 1990; Csink and Henikoff 1998).

The structures of yeast centromeres (both *Shizosaccharomyces pombe* and the point centromeres of *Saccharomyces cerevisiae*) have been derived; until recently, however, the sequence structure of centromeres of higher eukaryotes had been limited to gross sequence organization. The small, heterogeneous sequence structure of several rice centromeres was exploited to complete the first full characterization and contiguous assembly of centromeres in a higher eukaryote (Nagaki et al. 2004; Zhang et al. 2004). For example, the 1.65 Mb region of the centromere of chromosome 8 (*Cen8*) contains a delimited ~750 kb functional core with active genes and an enrichment of young, active centromeric retroelements (CRRs) but lacks long expanses of satellite arrays (Nagaki et al. 2004). It is hypothesized that this centromere is in an intermediate stage of its evolution, with the potential to garner long expanses of homogenous arrays as it ages (Nagaki et al. 2004; Chap. 6 in this book).

4.1.2 Functional and Epigenetic Components of the Centromere

While the sequences comprising the mammalian centromere have been difficult to capture, the protein organization has been well characterized. Residing at the inner kinetochore plate of active centromeres (Warburton et al. 1997; Sullivan and Karpen 2004), the centromere-specific protein CENP-A (also referred to as CID, cenH3) replaces histone H3 (Sullivan et al. 1994) and is interspersed with normal histone H3 containing nucleosomes (Sullivan and Karpen 2004; Lam et al. 2006). Although the method of CENP-A deposition at the centromere is unknown, it is localized exclusively to active centromeres and may not simply bind DNA, but may be recruited through an unknown epigenetic pathway to the kinetochore (Mellone and Allshire 2003). The centromere protein CENP-B localizes to the central domain of the mammalian centromere defined by the presence of centromeric heterochromatin. CENP-B binds a conserved DNA binding motif, known as the CENP-B box, within α-satellite centromeric DNA in humans (Masumoto et al. 1989). This 17 bp motif is highly conserved from human to Australian marsupials (Bulazel et al. 2006), yet the presence of CENP-B is not necessary to maintain kinetochore function in CENP-B knockout mouse cells (Hudson et al. 1998). However, CENP-B has been determined to be essential in the formation of de novo human centromeres (Ohzeki et al. 2002).

Several heterochromatin-specific proteins are also involved in centromere function and include H3 variants that have specific modifications to amino acid residues (Lachner and Jenuwein 2002; Elgin and Grewal 2003; Lachner et al. 2003). For example, trimethylation of lysine 9 (m3-H3K9) produces a modified histone found in the constitutive heterochromatin at centromeres (Rice et al. 2003). Based on the studies on yeast, it has been proposed that RNAs mediate the pairing of centromere-specific DNAs to chromodomain-like adaptor proteins, which in turn recruit histone methyltransferases (HMTases) that target the H3K9 residue for methylation. This interaction may be stabilized by the centromere-specific heterochromatin protein 1 (HP1) (Nakayama et al. 2001; Hall et al. 2002). The methylation of H3K9 also triggers DNA methylation of CpG residues in centromeres (Fuks et al. 2003).

Chromatin immunoprecipitation (ChIP) has been instrumental in isolating specific DNA that binds centromere proteins in mammals. Studies using this technique in human artificial chromosomes demonstrated that different centromere chromatin domains are neither satellite sequence dependent nor specific to large, homogenous arrays of centromeric satellites (Lam et al. 2006). Instead, specific regional proteins, such as CENP-A and H3K9, are dynamic: they target centromere domains in a nonsequence-dependent manner and can spread across non-satellite DNA. Thus, a conserved centromere *organization* in eukaryotes may be more important in regulating centromere function than the satellite sequences themselves (Partridge et al. 2000; Pidoux and Allshire 2004; Sullivan and Karpen 2004; Lam et al. 2006). However, the lack of *complete* molecular maps for any mammalian centromere, and thus the inability to characterize the overall landscape and spatial organization of

the elements at the mammalian centromere core, has limited our understanding of the epigenetic framework of the mammalian centromere. Recently, a mammalian model for centromere structure and organization with respect to the kinetochore-delimited core region has been identified: the Australian marsupial, *Macropus eugenii* (the tammar wallaby).

4.2 Marsupial Models for Studying Centromere Function and Evolution

Having last shared a common ancestor with eutherian mammals ~166 million years ago (mya) (Bininda-Emonds et al. 2007), marsupials are ideally situated as comparative models for eutherians and have afforded numerous insights into mammalian physiology, ecology, evolution, and genetics (Renfree 2006). Within mammals, there are three extant infraclasses: Eutherians, Marsupials, and Monotremes. The latter group last shared a common ancestor with that of both eutherian and marsupial mammals ~180 mya, placing marsupials in a unique position in which to infer their ancestral states (Fig. 4.1).

Studies of chromosome evolution in mammals have focused heavily on the evolution of conserved syntenic, gene-rich domains. It is also apparent that the centromere plays an equally important role in chromosome evolution through its involvement in fissions, centric fusions, translocations, inversions, and centric shifts (also referred to as centromere emergence/repositioning). Several mammalian systems show a dramatic level of karyotypic diversity (see Eldridge and Close 1993; Wang 2000, for examples), frequently involving centromere-associated rearrangements. The central question remains: could karyotypic diversity be *driven* by centromere-associated changes? Henikoff and colleagues (Henikoff et al. 2001; Henikoff and Malik 2002; Malik and Henikoff 2002) have proposed that the centromere is a selfish entity, citing as evidence the extremely rapid evolution of centromeric satellites that is tracked by positive directional selection of the centromeric histone H3, cenH3 (Malik and Henikoff 2001; Talbert et al. 2002; Chap. 2 in this book), in several

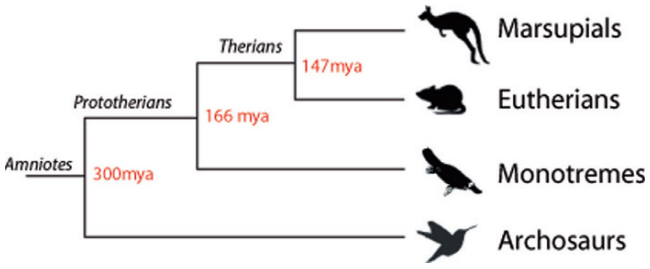


Fig. 4.1 Phylogeny of extant mammalian infraclasses. Divergence date approximations taken from Bininda-Emonds et al. (2007)

organisms. The rapid evolution is envisioned as an arms race between selfish DNA elements capable of distorting chromosome segregation (in female meiosis) in their favor, while cenH3 is selected to maintain equal segregation. Thus, it has been proposed that centromeres, acting in their "own interest," facilitate the creation of chromosomal rearrangements that lead to reproductive isolation and, ultimately, the emergence of new species (O'Neill et al. 2004; Metcalfe et al. 2007). The rapid karyotypic evolution within marsupial mammals has afforded an exciting and unique model system in which to study centromere emergence, evolution, and function.

4.2.1 Marsupial Karyotypic Diversity

Marsupial chromosomes are among the most widely studied of any mammalian group, with over 70% of the known ~334 marsupial species karyotyped (Hayman 1977, 1990). Central to the description of this karyotypic diversity is the involvement of the centromere, either through its location on the chromosome or its involvement in fissions, translocations, fusions, and shifts within marsupial genomes.

Among marsupials, macropodines have the most extensively studied and well-characterized karyotypes in terms of G-banding, chromosome rearrangements, and homologies. Interestingly, they also carry the most diverse array of karyotypes, with diploid numbers ranging from $2n = 10,11$ in *Wallabia bicolor* to $2n = 22$ found in several species and considered to be the ancestral karyotype for this subfamily (Rofe 1979; Hayman 1990). The macropodines can be subdivided into three groups, each representing different rates of karyotypic evolution.

The first group contains all the species within the genus *Thylogale* (pademelons). All the species within this group retain the plesiomorphic macropodine karyotype, $2n = 22$, and have undergone an apparent slow rate of karyotypic evolution.

The second group contains species within the karyotypically diverse genus *Macropus* (kangaroos and wallabies) as well as the monotypic genus *Wallabia* (the swamp wallaby). These species harbor karyotypic differences attributed mainly to centric fusions. Most species carry a morphologically similar diploid complement of $2n = 16$, the chromosomes of which often represent different suites of fusions (Hayman 1990). For example, *M. giganteus* and *M. eugenii* look karyotypically identical using light microscopy; however, each carries a high level of intra-chromosomal variation with respect to the proposed macropodine ancestral $2n = 22$ karyotype (Rofe 1979; Bulazel et al. 2007).

The third group within macropodines is comprised solely of the genus *Petrogale* (rock wallabies). Having undergone a recent and rapid explosion of chromosomal evolution (Eldridge and Close 1993), all 21 taxa within this genus exhibit distinct chromosomal complements, with the exception of two sub-species (*Petrogale xanthopus xanthopus* and *Petrogale xanthopus celeris*) (Sharman et al. 1990). Centric fusions, centric shifts, and inversions are characteristics of the majority of *Petrogale* taxa (see Eldridge and Close 1993 for a review). New centromere emergence occurs at a high frequency within this group; however, inversions are not

responsible for the apparent mobility of the centromeres of this genus, rather the centromere location has shifted relative to the ancestral state for any particular chromosome (Eldridge and Close 1993).

Nineteen chromosome segments comprise the marsupial karyotype (North and South American as well as Australian) and have been conserved as large syntenic blocks within all species examined thus far (Rens et al. 1999, 2003; O'Neill et al. 2004). Using phylogenetic and karyotypic approaches, the presumed ancestral Marsupial karyotype has been derived with respect to the 19 conserved blocks (Rens et al. 1999, 2003). From this karyotype, the "shuffling" of conserved blocks through rearrangement can be seen in many extant lineages (see examples Fig. 4.2). Many of the breaks between these blocks have undergone convergent breakpoint reuse in karyotypic rearrangements across disparate marsupial lineages (Rens et al. 2003). Several of these rearrangements involved centromere repositioning within a single chromosome, resulting in a large number of potential latent centromeres concentrated at breaks between conserved chromosome segments within several marsupial lineages (Ferreri et al. 2005).

For example, tracing the phylogenetic history of marsupial conserved segments 13 (C13) and 14 (C14) on chromosome 2 through cross-species reciprocal chromosome painting (Rens et al. 2003) and G-band analyses (Rofe 1979; Hayman 1990;

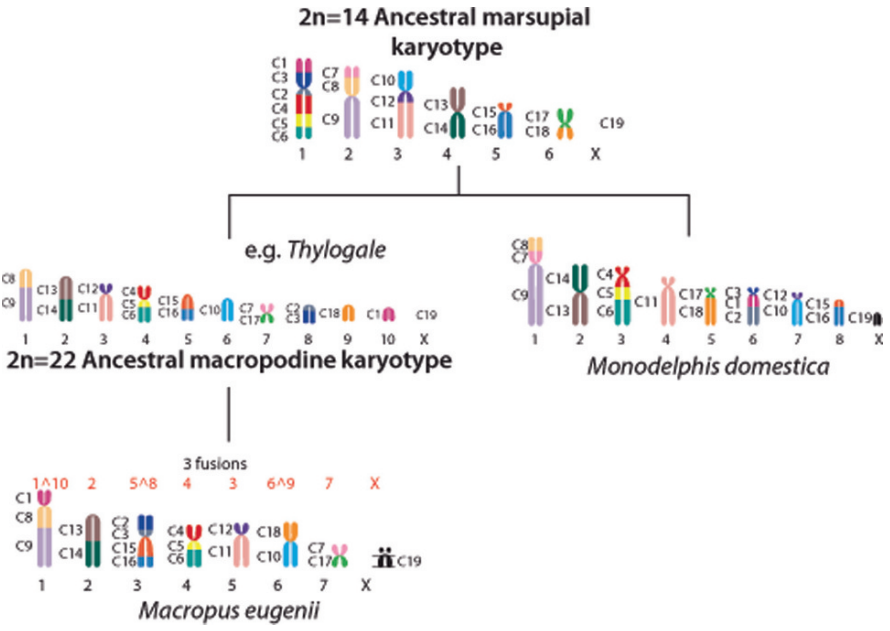


Fig. 4.2 Derivation of extant and ancestral karyotypes within marsupials with respect to the 19 conserved chromosome segments. Each segment has been color-coded. Extant karyotypes are represented by *M. eugenii* (tammar) and *Monodelphis domestica* (South American opossum). Hypothetic ancestral karyotypes for Macropodinae and all marsupial lineages are indicated. The rearrangements required to generate the tammar karyotype are indicated

Svartman and Vianna-Morgante 1999) revealed that these segments have experienced fissions into two separate chromosomes in two divergent lineages who last shared an ancestor >65 mya (*Didelphis marsupialis*, the North American opossum, a member of the American marsupials, and *Trichosurus vulpecula*, the brush-tailed possum, an Australian marsupial). As shown in Fig. 4.3, these two species have formed new centromeres as part of this fission event. In the case of *T. vulpecula*, the C14 fragment was also involved in a fusion without an apparent inversion of material, indicating that the centromere formed through the fission event on C14 was silenced and may be retained in latent form.

Within the karyotype of the ancestor to the Macropodidae, the family of kangaroos, wallabies and potoroos, including the wallaby *M. eugenii*, there has been a centric shift of this chromosome from a metacentric form to an acrocentric form, again in the absence of inversions. This shift is shared in several lineages, including all Macropodinae (kangaroos and wallabies, including *Macropus* spp.) and Potoroinae (potoroos and bettongs, including *Aepyprymnus* spp. which carries another shift) (Rens et al. 2003; O'Neill et al. 2004). Thus, in the context of this phylogenetic history, the syntenic block C13 within *M. eugenii* harbors two types of centromeres. The first is the active centromere. This is the site of spindle attachment and kinetochore assembly (black circles, Fig. 4.3). The second is the latent centromere found between C13 and C14 in *M. eugenii*, where the centric shift occurred while leaving behind centromere sequences (grey circle, Fig. 4.3; Ferreri et al. 2005; and see below). Repositioning events such as this example are a recurring feature of marsupial karyotypic evolution, providing an ideal model for understanding ectopic centromere emergence in an evolutionary context.

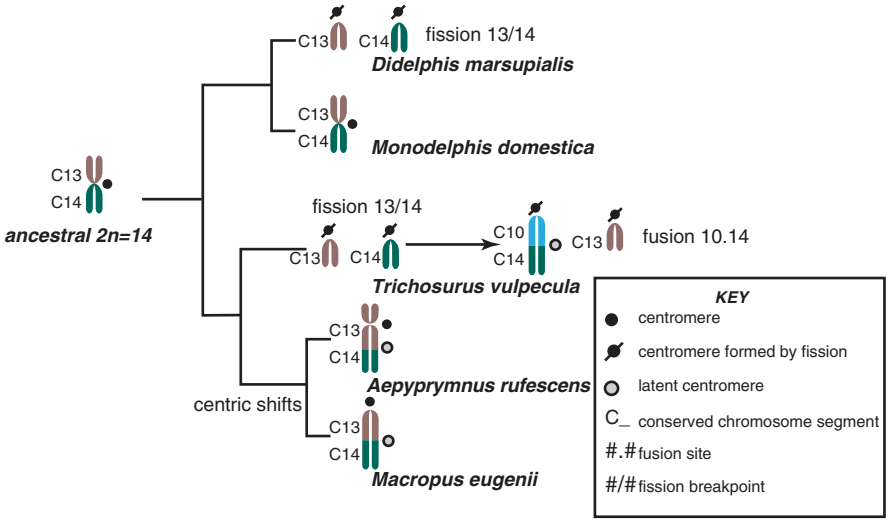


Fig. 4.3 Phylogenetic tree of *Marsupialia* species with informative chromosome rearrangements for chromosome 2, indicating the evolutionary path of conserved segments C13 and C14. The centromere is shown to the right of metacentric chromosomes and above acrocentric chromosomes. Latent centromeres are indicated, as those that are the result of fission (as per KEY in inset)

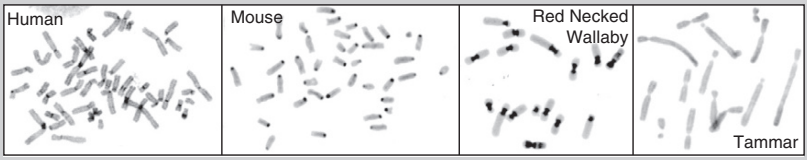
4.2.2 Latent Centromeres in Marsupials

The ectopic emergence of a de novo centromere most frequently occurs to provide mitotic stability to otherwise acentric chromosome fragments resulting from rearrangement (Amor and Choo 2002; Warburton 2004). In similar manner, ectopic centromeres can appear on otherwise normal chromosomes to create dicentrics. Approximately 70 described cases of neocentromeres have been identified on 19 human chromosomes (Warburton 2004). Almost 10% of these cases are meiotically stable and heritable (Knegt et al. 2003; Amor et al. 2004). Three clear "hot spots" for neocentromeres have been identified within the human karyotype (3q26-qter, 13q21-32, and 15q24-26) (Amor and Choo 2002), implying a nonrandom mechanism for their appearance. This has implications not only for the role neocentromeres play in human genetic disease but also for their role in creating karyotypic diversity involving repositioning of a centromere.

Latent sites giving rise to dicentrics and neocentromeres lack the satellite sequence features characteristic of normal chromosomes (Sullivan and Willard 1998; Barry et al. 1999; Lo et al. 2001a, b; Alonso et al. 2003), suggesting that satellite DNA is not necessary for the demarcation of a new centromere location. An epigenetic mechanism for the repatterning of a segment of chromatin to perform as a competent site of kinetochore attachment and assembly has, therefore, been hypothesized as the priming event for centromere emergence (Choo 1997b; du Sart et al. 1997). Under its initial description, this "latent centromere hypothesis" relies on the presence of a centromere-specific sequence at the site of imminent centromere formation. Recently, this hypothesis has been modified to suggest that there may be latent chromatin and/or genomic structures that act as a mark for centromere formation (Ventura et al. 2004).

Using FISH with BAC probes, labeling several human chromosomes, and in silico analyses of the BAC sequences, Ventura et al. (2003) identified a putative latent centromere in 15q25. This centromere was inactivated at the time of the fission event that resulted in chromosomes 14 and 15 and the emergence of two new centromeres. This ancestral location coincides with neocentromere formation in 15q24-26 in at least two human cases, further supporting the latent centromere hypothesis. We have applied a similar approach for studying the relationship between the breakpoints conserved between the 19 chromosome blocks that define marsupial karyotypes, the evolution of centromere sequences and resident retroelements. Through previous work on interspecific hybrids, we identified a conserved retroelement, KERV, that is found within the centromeres of two of our model species, *M. eugenii* and *M. rufogriseus* (the red necked wallaby) (see Box. 1). Screening a *M. eugenii* BAC library with a portion of KERV, we mapped 48 KERV-positive BACs to *M. eugenii* metaphase chromosomes. While expecting centromere localization, we were surprised to find that these BACs map to breakpoints between the 19 conserved chromosome blocks as well as centromeres and telomeres (Fig. 4.4). Some of the BAC locations (red arrows, Fig. 4.4) were not previously identified as breakpoints, but phylogenetic inference has shown that these are ancestral centromere locations

Box 1 Marsupial models



Our efforts to isolate centromere-specific sequences initially focused on the macropodine (kangaroos and wallabies) species *Macropus rufogriseus banksianus*. This species has a $2n = 16$ karyotype with an identical complement to that of *M. eugenii*. However, each chromosome of *M. rufogriseus* carries unusually extensive constitutive heterochromatin at the centromeres compared to other mammalian species, including its sister species, as determined by C-banding (Rofe 1979; Lowry et al. 1994). The size of these regions allowed for easier manipulation by microdissection for isolation of centromere DNAs

within marsupials. For example, one of these locations, within C10 on chromosome 1, is a known break of synteny between *M. eugenii* and *M. domestica* (the South American opossum, Deakin, personal communication), with whom *M. eugenii* last shared an ancestor for ~65 mya.

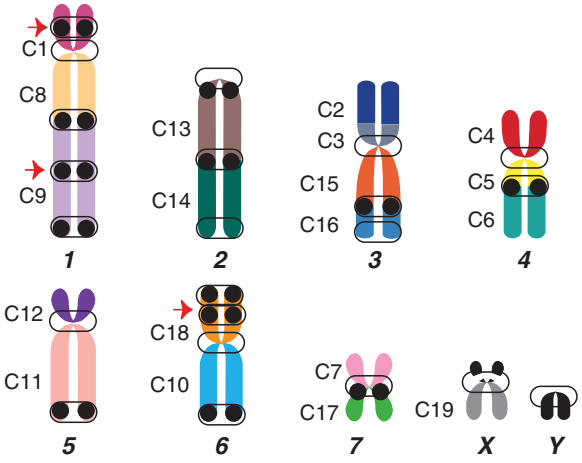


Fig. 4.4 Map of KERV locations within the *M. eugenii* karyotype overlaid on an ideogram showing the 19 chromosome segments conserved in marsupials (colored and labeled to the left of each segment) along each chromosome (listed to the bottom). The locations of KERV sequences identified by Fluorescence in situ hybridization (FISH) are indicated in *oval* and those identified by BAC mapping are indicated by *dots*. Arrows highlight three KERV locations that are centromeres in another marsupial species yet are not considered breaks between conserved chromosome blocks

Retention of specific centromere sequences at evolutionary breakpoints provides an intriguing correlation between retroelements and the reshuffling of chromosome blocks in marsupials. Deactivation or reactivation of a centromere (from a latent to an active centromere in the case of the former and vice versa in the case of the latter) may be facilitated by increased retroelement activity such as accompanying genome instability (see Box 2 for a discussion of hybridization-induced instability at centromeres).

4.2.3 Identification and Functional Characterization of Centromeric Satellites in Macropus

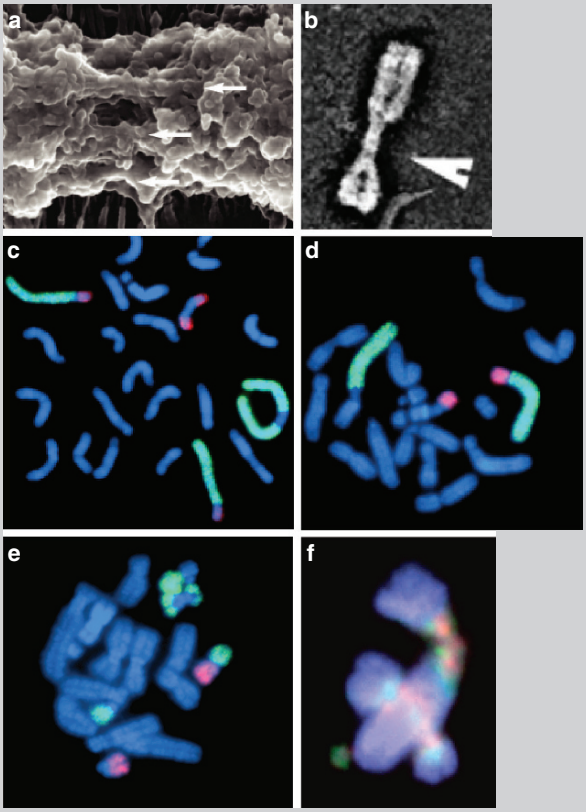
Initial efforts to isolate other centromere-specific sequences again focused on *M. rufogriseus*. Using a combination of microdissection and microcloning, the centromere of the X chromosome from this species was isolated. Within this centro-

Box 2 Analysis of hybrid genomes.

Further evidence for the correlation between centromere dynamics and karyotypic diversity in marsupials has been found in the genomes of interspecific hybrids within the *Macropus* genus. Several dysgenic hybrids display karyotypic aberrations almost exclusively associated with centromeric abnormalities, including translocations and amplifications (O'Neill et al. 1998, 2001). Detailed analysis of several hybrids from different interspecific crosses has shown instabilities linked to the retroelement KERV, attributed to a significant copy-number increase of this sequence in the centromere (O'Neill et al. 1998; Metcalfe et al. 2007). Recent research has shown that this centromeric amplification also associates with fusion and fission events, as well as knob-formation, a potentially meiotically driven element (Rhoades and Dempsey 1966). Thus it appears that the centromere, or at least centromere-associated sequences, may have played a pivotal role in chromosome restructuring and centromere repositioning in macropodines.

Examination of several marsupial interspecific hybrids has suggested that chromatin remodeling and genomic rearrangements are restricted to the centromere (O'Neill et al. 1998, 2001; Metcalfe et al. 2007). In particular, *Macropus rufogriseus* x *Macropus agilis* hybrid chromosomes are typified by centromere abnormalities and rearrangements involving the centromere of the maternal complement (*M. rufogriseus*). Large blocks of heterochromatin surrounding the centromere characterize *M. rufogriseus* chromosomes, whereas the centromeres of the paternal species, *M. agilis,* consist of very little heterochromatin. The centromeres of both species are comprised of two predominate sequences, the α-like satellite sat23 and the endogenous retrovirus KERV, but differ in relative abundance of these sequences (Bulazel et al. 2006).

Box 2 (continued)



M. rufogriseus x *M. agilis* hybrids demonstrate an increase in both sat23 and KERV copy number and abnormally extended maternal chromosomes (Metcalfe et al. 2007) (**a**) indicating there is an amplification of KERV and sat23 at the centromere. In conjunction with this, scanning electron microscopy indicates that the hybrid centromeres have an increase in DNA content at the centromere, with an uneven distribution of DNA throughout the hybrid centromere (**b**) as compared to the normal maternal centromere (Metcalfe et al. 2007). Concomitant with these changes to maternally derived centromeres is a markedly higher incidence of centromere-limited chromosome rearrangements, including (**c**) isochromosomes, (**d**) whole arm reciprocal translocations, (**e**) fissions, and (**f**) minichromosomes

mere DNA library, we identified a sequence class, sat23, that is a 178 bp repeat with long-range periodicity that contains the CENP-B 17 bp DNA binding domain, actively binds CENP-B in vitro and in vivo, and localizes to the centromeric region of every chromosome (Bulazel et al. 2006, 2007). Thus, sat23 represents the

alphoid-like satellite of this species. Two other sequences were identified within this library of cenDNA: sat1, a sequence restricted to the centromeres of the sex chromosomes of this species, and sat29, a sequence shared between the centromeres of the sex chromosomes and chromosome 2.

Using a comparative phylogenetic approach examining the conservation of these satellite classes across *Macropus* species, we uncovered a remarkable contradiction to the observation that satellites evolve rapidly; the satellite sat23 represents the major satellite component of most species within the *Macropus* lineage. The only exceptions to this are *M. giganteus* (the grey kangaroo) and *Wallabia bicolor* (the swamp wallaby). These species no longer carry this sequence as its predominant satellite; instead, the centromeres of the former species carry sat1 as its predominant centromeric satellite and the centromere satellites of the latter species are unknown.

The "true" phylogenetic history of *Macropus* species, determined by a combination of nuclear and mtDNA Baysian analyses, was compared to a phylogenetic tree derived from the most parsimonious relationships of these species determined solely by chromosome segments (Bulazel et al. 2007). This comparison shows that these two phylogenetic trees are discordant (Fig. 4.5). While the nuclear/mtDNA tree is clearly an accurate assessment of phylogenetic relationships of these species, the power of this comparison lies in the identification of breakpoint reuse within this group of mammals. In other words, the karyotypically similar species are not necessarily phylogenetic sister-taxa; rather, they have derived similar karyotypes through the reuse of specific breakpoints.

Mapping the satellite data for the aforementioned sequences back onto the "true" phylogenetic tree of these species provided some insight into the conservation of

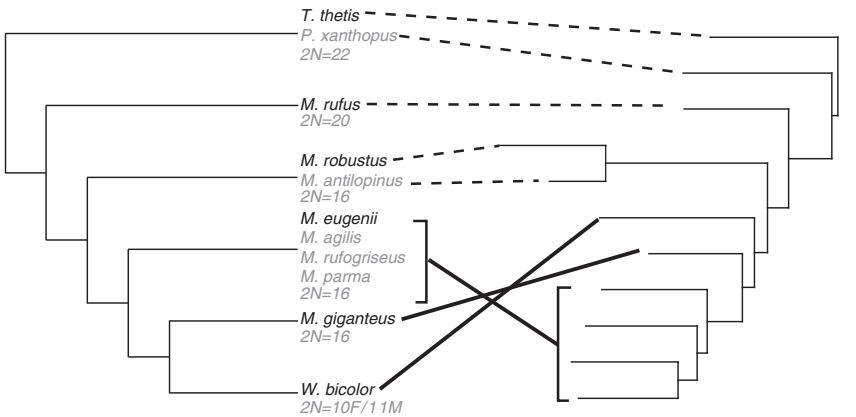


Fig. 4.5 Comparison of the *Macropus* phylogenetic tree derived from nuclear/mtDNA sequences employing Baysian approaches (*left*) compared to the *Macropus* phylogenetic tree derived from an analysis of the conserved chromosome blocks employing the GRIMM algorithm (*right*). The lineages for which there is tree topology agreement are indicated with *dashed lines* while the lineages for which these two trees are discordant in topology are shown with *solid lines*

satellite sequences in distantly related taxa within this group of mammals. It appears that the reuse of breakpoints within *Macropus* is restricted to specific chromosomal segments (C1, C2, C8, C10, C15, C18) and that the conservation of satellites is coincident with the reuse of these segments (see Fig. 4.6 for an example in *M. robustus*). Thus, convergent breakpoint reuse may be the mechanism by which these sequences remain at specific centromeres (Bulazel et al. 2007).

4.2.4 *Centromere Size and Gross Organization within* M. eugenii

The large size of the centromere domains of *M. rufogriseus* precludes further long-range sequence analysis. Therefore, the conservation of sequences identified in *M. rufogriseus* was investigated within another $2n = 16$ macropodine species, *M. eugenii*, the tammar wallaby. As a model for centromere research, the tammar is markedly different from *M. rufogriseus*: its centromeres are extremely small and the constitutive heterochromatin content is so low as to be undetectable by C-banding (see Box 1). Despite the difference in overall quantity of centromeric DNA between

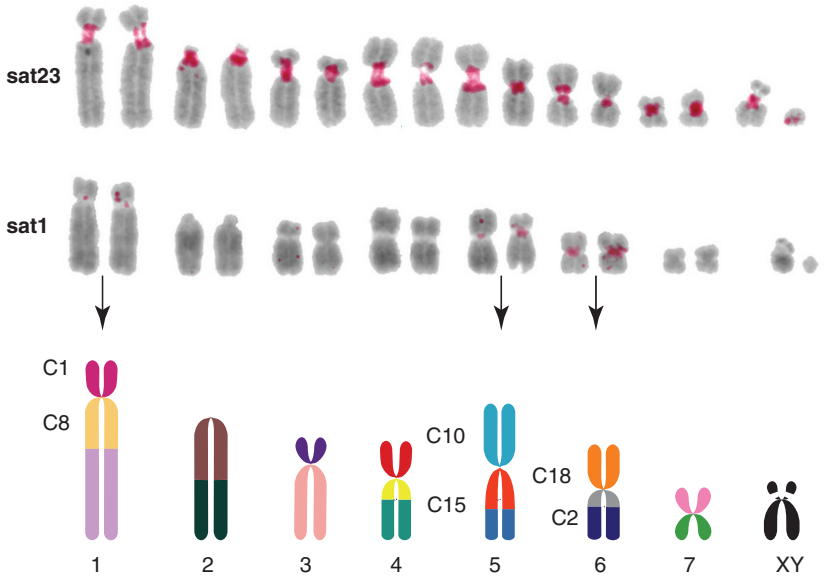


Fig. 4.6 FISH mapping of satellites (*red*) sat23 (*top*) and sat1 (*middle*) to metaphase chromosomes of *M. robustus* (the wallaroo). Arrows indicate the chromosomes (1, 5, and 6) that have retained sat1 sequences through breakpoint reuse of the conserved segments (*in bold*) that are reused in multiple *Macropus* lineages. The ideogram for *M. robustus* with respect to the 19 conserved chromosome segments is shown at the bottom

these two species, sat23 is also found at the centromeres of every chromosome in tammar (Fig. 4.7).

Like sat23, the retroelement KERV is found concentrated at tammar wallaby centromeres (Ferreri et al. 2004; Fig. 4.7), and FISH demonstrates that both sequences occupy the same centromeric domains. Fiber FISH on single DNA fibers indicates that the two sequences do not occupy separate, juxtaposed blocks but are instead interspersed with one another throughout the centromere. Significantly, this overall hybridization pattern is not concordant with the pattern of blocks of tandemly arrayed satellites observed in mouse and human, but is most similar to the hybridization pattern observed for the Cent-O satellite and CRRs found in rice centromeres (Fig. 4.7). The similarity between the organization of the small centromeres within a mammal and a plant, two disparate lineages, support the hypothesis that a conserved centromere structure exists for higher eukaryotes. Under a model for centromere structure where the core contains retroelements and satellites interspersed with one another, the accumulation of large tracts of satellites surrounding this core occurs after fixation (and likely over long periods of chromosomal stability) of the newly formed centromere within a population (Table 4.1).

The second striking observation from fiber FISH experiments is that the centromeres of tammar are also similar to several rice centromeres in overall size. On the basis of kb/micron calibration, the average length of centromeres across all chromosomes within tammar is 420.2 ± 14.4 kb (Carone et al. 2009). This was confirmed by measurements of centromere length in fiber FISH and immunofluorescence (IF)

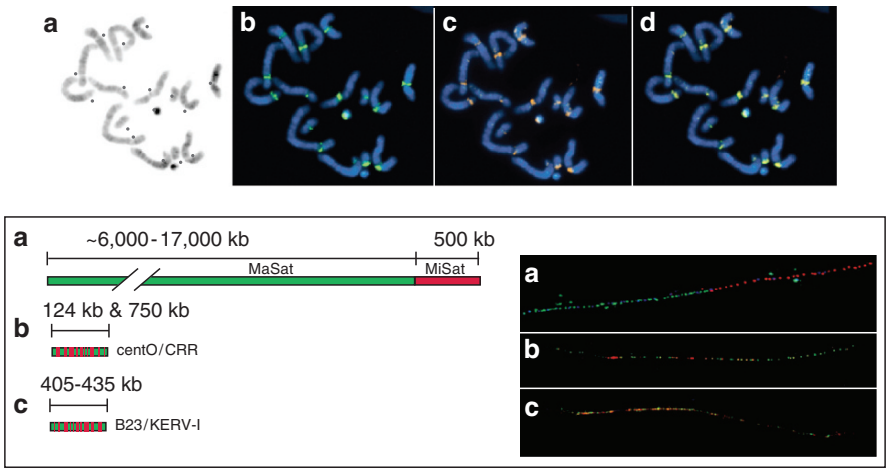


Fig. 4.7 FISH with centromere sequences to metaphase chromosomes (DAPI-stained Blue) of *M. eugenii*. *Top*: (**a**) Inverted DAPI image with the centromeres indicated with a red spot; (**b**) sat23 (*green*), (**c**) KERV (*orange*), (**d**) merged image. *Bottom*: Composite image illustrating the structural differences between (**a**) mouse centromeres (Garagna et al. 2002; Kuznetsova et al. 2006) and (**b**) rice *CEN4* and *CEN8*, respectively (Cheng et al. 2002) and (**c**) *M. eugenii* centromeres using fiber FISH mapping. To the *left* is a representation of probe order and overall centromere size and to the *right* is the corresponding FISH images

Table 4.1 Summary of repeat and RNA transcript data known for several eukaryotic model species, discussed earlier

Organism	Centromere repeats	Centromere RNAs	Small RNAs	dsRNAs	Centromere retroelements
Yeast[a]	Otr repeats	Yes	siRNAs	Yes	no
Rice[b]	CentO	Yes	siRNAs	Yes	CRR
Maize	CentC	Yes	unknown	Yes	CRM
Tammar	sat23	Yes	Yes	Yes	KERV
Mouse	*major, minor*	Yes	inferred	Yes	KERV[c]
Human	*alpha, gamma, satIII*	Yes	inferred	Yes	LINE-1

Centromere repeats in italics are only a subset of the specific satellite sequences known and are carried as tandem arrays

[a] S. pombe

[b] Represented by a subset of rice chromosomes

[c] O'Neill, unpublished

experiments using CREST sera (Carone et al. 2009), containing antibodies for kinetochore proteins CENP-A, B, and C and supported by the lack of C-band positive material at these centromeres (see Box 1), indicating a microscopically undetectable amount of heterochromatin in this species. This centromere size is *notably* smaller than the 2–5 Mb centromeres/pericentromeres of human and 6–20 Mb centromeres/pericentromeres of mouse (Choo 1997a; Fig. 4.7) and provides a model system in which to study centromere structure and function. Current studies are now focused on the functional components of the tammar centromere, including the involvement of RNAs in centromere maintenance.

4.3 Noncoding RNA and the Centromere

Centromeres have long been thought to comprise noncoding and transcriptionally inactive DNA. However, recent evidence suggests that eukaryotic centromeres produce a variety of transcripts. The transcription of satellites has been observed in numerous eukaryotic species across a broad range of phyla, from yeast to human (Diaz et al. 1981; Miyahara et al. 1985; Epstein et al. 1986; Wu et al. 1986; Bonaccorsi et al. 1990; Belyaeva et al. 1992; Rudert et al. 1995; Rouleux-Bonnin et al. 1996; Renault et al. 1999; Lachner and Jenuwein 2002; Volpe et al. 2002, 2003; Lehnertz et al. 2003; Li and Kirby 2003; Fukagawa et al. 2004; Topp et al. 2004; Bouzinba-Segard et al. 2006; Lee et al. 2006). The wide-spread conservation of satellite transcription is consistent with a conserved regulatory role for these transcripts in gene regulation or chromatin modification (Ugarkovic 2005). These transcripts may function in one of the three ways: (1) They may facilitate post-transcriptional gene regulation (Li and Kirby 2003), potentially through the RNA-induced silencing complex (RISC). In this pathway, double stranded (ds) RNAs are cleaved into short interfering RNAs (siRNAs, 21 nucleotide

double stranded RNAs) that, upon association with RISC, mediate native mRNA inactivation (Hammond et al. 2000). (2) They may participate in the RNA-induced transcriptional silencing complex (RITS), a pathway in which siRNAs are involved in heterochromatin recruitment (Volpe et al. 2002, 2003). (3) Alternatively, in a manner analogous to the *Xist* transcript in mammalian X-inactivation, they may recruit heterochromatin assembly factors such as histone deacetylases, SET domain proteins, and Polycomb group proteins (Heard 2005). Although the mechanisms are unknown, evidence that satellite transcripts participate in heterochromatin assembly and/or nucleosome recruitment at centromeres is accumulating.

In *Shizosaccharomyces pombe* centromeres, dsRNAs transcribed from the *dh* and *dg* repeats in the pericentric *otr* region produce siRNAs that are bound to the RITS complex and bring about H3 lysine-9 methylation through the RNA interference pathway (RNAi) (Volpe et al. 2002, 2003). In maize, transcripts have been identified from both strands of the 156 bp CentC centromere-specific repeat as well as the centromere-specific CRM retroelement, each of which coimmunoprecipitates with the CENP-A antibody. Although no siRNAs were found in this study (Topp et al. 2004), siRNAs have been identified for CentO repeats, the analogous centromere-specific repeat in rice (Lee et al. 2006), indicating that the RNAi pathway may be involved in centromere transcript processing in plants. Thus, a complex interaction of RNAs, modified histones, and DNA define the genomic locations that act as centromeres. Recent work in mouse, human, and our work in tammar suggests that this may also be true of mammalian centromeres.

Obliteration of dsRNA in mouse results in the loss of centromere foci in interphase nuclei (Maison et al. 2002). Mouse cells null for *dicer*, the gene encoding the enzyme responsible for cleaving dsRNA into siRNAs, show a similar centromere defect (Peters et al. 2001; Kanellopoulou et al. 2005), implicating an RNA silencing pathway in centromere function in mammalian cells through dsRNA processing. Fukagawa et al. (2004) used human–chicken somatic cell hybrids to demonstrate that *dicer* conditional loss of function mutant cells lack centromeric heterochromatin and exhibit an accumulation of centromere satellite transcripts, implicating the need for *dicer* to cleave them into smaller RNAs. From these studies, it has been proposed that centromere satellite transcripts have a role in kinetochore assembly in mammals through kinetochore demarcation and heterochromatin establishment (Fukagawa et al. 2004; White and Allshire 2004).

The transcription of centromere sequences appears to be under strict regulation in human and mouse cells. Stresses, such as heat shock, nutrient deficiency, apoptosis, and chemical shock result in genetic instability that ultimately leads to aneuploidy, loss of sister chromatid cohesion, and abnormal chromosome segregation. These defects are directly correlated with aberrant transcription of centromere satellites. In mouse, 120 nt transcripts for the minor satellite accumulate under stress conditions that ultimately lead to abnormal centromere function (Bouzinba-Segard et al. 2006). Similar aberrant transcript accumulation has been found for satellite III (satIII) satellites in human cells under stress conditions (Valgardsdottir et al. 2005). Based

on these studies, and the *dicer* deficient cell assays, it has been proposed that the accumulation of these transcripts results from improper RNA processing of larger transcripts, resulting in a reduction of small RNAs that participate in the recruitment of specific histones critical for centromere functioning.

In mammals, however, the large size of mammalian centromeres and the presence of large tracts of repetitive DNA within them have limited studies on the role of small RNA transcription in centromere function. Previous studies have failed to identify the class of small RNAs produced from mammalian centromeres, their native transcript forms, and the regional boundaries of transcriptional activity. Most importantly, the mechanism through which transcription of these satellite sequences is promoted is unknown (White and Allshire 2004). It has been proposed that transcriptional control through retroelements may facilitate the satellite sequence transcription observed in a broad range of vertebrate species (Diaz et al. 1981; Ugarkovic 2005; see Sect. 4.3.2).

4.3.1 The Role of snRNA in Marsupial Centromeres

We have previously highlighted the localization of a retroelement, KERV, to centromeres of Macropodines (see Sect. 4.2.2). KERV is an endogenous retrovirus (O'Neill et al. 1998) characterized by open reading frames for *gag, pro*, and *pol* bounded by two identical long-terminal repeats (LTRs) and is found in all Macropodine lineages (Ferreri et al. 2004). The striking similarity between the interspersed arrangement of retroelements and centromeric satellites in rice (Cheng et al. 2002) and maize (Jin et al. 2004) and the interspersed arrangement of KERV and the centromeric satellite sat23 (Carone et al. 2009; see Sect. 4.2.4) concomitant with the discovery of siRNA emanating from CentO satellite transcripts in rice (Lee et al. 2006) compelled an investigation into the role of transcription and small noncoding RNA in macropodine centromeres.

To this end, RNA depletion experiments followed by immunocytochemistry localization of centromere and heterochromatin proteins indicated that RNA is necessary for the recruitment of centromere (CENP-A and CENP-B) and heterochromatin (tri-methyl H3K9) proteins (Carone et al. 2009). Further investigation into the RNA species involved in this association and the transcripts produced from known centromeric sequences and, in particular small noncoding RNA, indicated that small RNA transcripts produced from *M. eugenii* centromeres are not in the size range of siRNA (21–23 nt) as seen for plant and yeast satellite sequences. In contrast, the small RNA produced from the wallaby centromeres are 34–42 nt, a previously unknown size class termed crasiRNAs (centromere repeat associated small interacting RNAs) (Carone et al. 2009). Furthermore, we propose that the production of crasiRNAs occurs via a dsRNA intermediate facilitated by the known bidirectional promoter capability of the KERV LTR (Carone et al. 2009) (far left, Fig. 4.8). We hypothesize that these small RNAs are tightly linked to retroelement activity and are integral to centromere functioning.

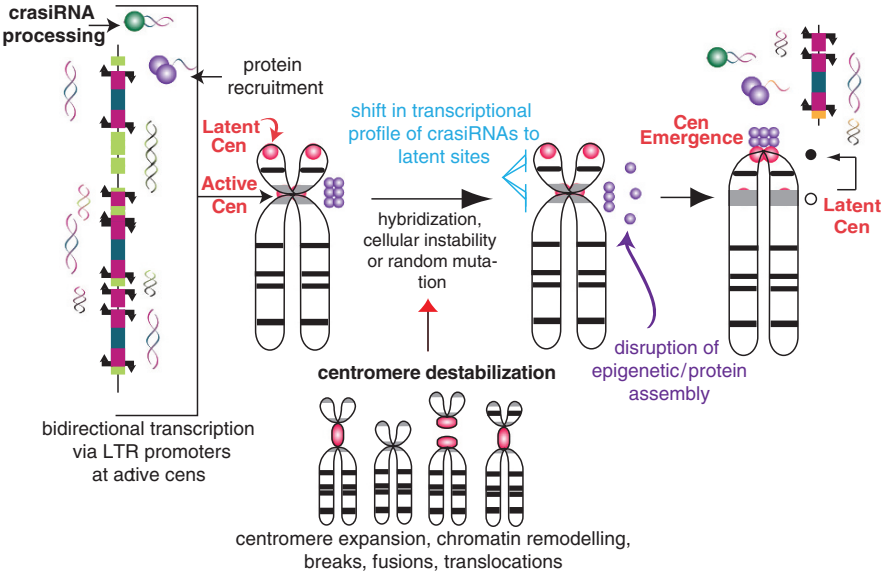


Fig. 4.8 Proposed model of centromere transcription and its role in centromere emergence (i.e., reactivation of a latent site). Transcription of centromere sequences at active centromeres (cen) is mediated by the bidirectional promoters of the LTRs (magenta boxes to left and right and red circles on chromosomes). Green boxes on the left represent sat23 while orange boxes on the right represent newly derived satellites, blue boxes are the internal portion of KERV, spaces represent other, as yet, unidentified sequences. Double stranded RNA transcripts are shown and are processed into crasiRNAs via an unknown pathway (*green*). Putative recruitment of cen proteins or epigenetic modifiers (purple) to active centromeres is mediated by single stranded crasiRNAs. Destabilization of the centromere, leading to translocations, fusions, fissions, and chromatin remodeling (shown at bottom as observed in hybrids reported herein), results in a shift of crasiRNA transcription to previously seeded KERV locations (latent cen)

4.3.2 Retroelements: An Integral and Functional Component of Centromeres?

Dawe (2003) and Wong and Choo (2004) have hypothesized that retroelements and their associated machinery may be integral to centromere functioning based upon three different lines of evidence. First, in plants some transposable elements have a genomic distribution restricted to the centromere. The centromere-specific retroelements in rice (CRR) are of the Ty3/gypsy class, map exclusively to centromeres (Cheng et al. 2002), and are strikingly dense in the kinetochore region of the centromere (Nagaki et al. 2004). Centromere retroelements (CRs) in both maize and rice associate preferentially with CENP-A (Zhong et al. 2002; Nagaki et al. 2004, 2005). Similar retrotransposon specificity for centromeres has been identified in many other plant species, including grasses, wheat and rye, and beet species (reviewed in Jiang et al. 2003). Interestingly, the LTRs (long terminal repeats) of the CRs of rice, barley, and

maize share significant sequence identity (Nagaki et al. 2003, 2005), implying constraint on their nucleotide sequence, contrary to both the centromere paradox and the commonly observed pattern of retroelement evolution.

LTRs act as strong promoters and are the primary means for an invading or mobilizing element to "highjack" the host's cellular machinery for self-replication. In this process, LTR promoters out-compete nearby native promoters for the same protein complexes, producing more retroelement RNAs (Coffin et al. 1997). The LTR promoters can retain their transcriptional potential once the sequence becomes integrated into the genome. As they age, these LTR sequences lose their ability to promote transcription through genetic drift and mutation caused by host defense mechanisms (Yoder et al. 1997). The retention of transcriptional machinery within the CR retroelement LTRs has led Jiang et al. (2003) to hypothesize that production of RNA transcripts by these LTRs facilitates the establishment of CENP-A domains in the demarcation of the active centromere.

Second, in several cases divergent repeat arrays within centromeres retain features of the retroelements from which they were derived (Wong and Choo 2004). For example, two clusters of tandem repeats, *ENSAT1* and *ENSAT2,* found in the pericentromere of *A. thaliana* chromosome IV share sequence similarity (72% and 79%, respectively) with the 5′ terminus of the *Atenspm2* transposon (Kapitonov and Jurka 1999). Thus, satellites found in centromere domains may be derived from retroelements, possibly through replication slippage, extensive deletion, or nonhomologous recombination.

Third, at least one centromere protein may have been derived from transposable element machinery. The amino acid sequence of CENP-B, a DNA-binding protein involved in the establishment of centric heterochromatin (see Sect. 4.1.2), shows significant similarity to *tigger*, a member of the TC1/mariner transposases (Kipling and Warburton 1997). The homologs of CENP-B in *S. pombe*, Cbh1 and Cbh2, both bind repeats found in the outermost pericentric block of DNA (*otr*) (Nakagawa et al. 2002). This interaction, likely mediated through siRNAs produced from specific repeats (*dg* and *dh*) in this block, is crucial for the establishment of H3K9 methylation at the centromere (Volpe et al. 2003). The coincidence of RNAs that are derived from retroelements found at plant centromeres and the association of RNAs and CENP-B homologs in the establishment of H3K9 methylation and constitutive heterochromatin formation in *S. pombe* further bolsters support for an integral role for transposable elements in the function of centromeres.

While the Dawe/Wong and Choo hypothesis has garnered robust support in plants (e.g., Zhong et al. 2002; Topp et al. 2004; Neumann et al. 2007), very little work has been done to test this theory directly in mammals. However, a recent study by Chueh et al. (2005) describes a positive correlation between neocentromere formation and transposable elements in humans, implicating LINE-1 in centromere initiation.

The observation of an interspersed arrangement of a centromeric satellite and a centromere-specific retroelement coupled with the evidence for the involvement of retroelements in centromeres provides the basis for a model of transcription of centromeric sequences in the tammar wallaby (Fig. 4.8). In this model, the strong bidirectional promoter capability of the KERV LTR produces long double-stranded

RNAs for both KERV and surrounding sequences (i.e., sat23) (Carone et al. 2009). This long dsRNA is then processed via an unknown mechanism into crasiRNAs, ~40 nt in length. The crasiRNAs are involved in the recruitment of heterochromatin and/or centromeric (kinetochore) proteins (Carone et al. 2009). The mechanism of this process may be similar to the recruitment of H3K9 via the RITS complex by siRNA emanating from *dg* and *dh* repeats in yeast (Volpe et al. 2002); however, the intermediate proteins involved in such a pathway are currently unknown. Interestingly, the observation of 40 nt snRNA associated with centromere proteins and sequences has also been reported in maize (Topp et al. 2004) and rice (Jin et al. 2004). Therefore, the production of snRNA, and in particular crasiRNAs, from centromeres and the involvement of small RNAs in recruiting centromere-specific proteins may be more conserved than previously thought.

Destabilization of centromeric chromatin states, perhaps through interspecies hybridization, cellular stress, or even random mutation, may shift the transcriptional activity of retroelements producing crasiRNAs from active centromere locations to previously seeded centromere locations (i.e., latent centromeres). It is unknown how this shift occurs and under what selection pressures fixation of such a centromere shift within a population might arise (Fig. 4. 8). It will be interesting to follow this field as we garner more insight into the components responsible for centromere protein deposition as well as the consequences of centromere mobility during species evolution.

References

Alonso A, Mahmood R, Li S, Cheung F, Yoda K, Warburton PE (2003) Genomic microarray analysis reveals distinct locations for the CENP-A binding domains in three human chromosome 13q32 neocentromeres. Hum Mol Genet 12:2711–2721

Amor DJ, Choo KH (2002) Neocentromeres role in human disease, evolution and, centromere study. Am J Hum Genet 71:695–714

Amor DJ, Bentley K, Ryan J, Perry J, Wong L, Slater H, Choo KH (2004) Human centromere repositioning in progress. Proc Natl Acad Sci USA 101:6542–6547

Barry AE, Howman EV, Cancilla MR, Saffery R, Choo KH (1999) Sequence analysis of an 80 kb human neocentromere. Hum Mol Genet 8:217–227

Belyaeva TA, Vishnivetsky PN, Potapov VA, Zhelezova AI, Romashchenko AG (1992) Species and tissue-specific transcription of complex, highly repeated satellite-like Bsp elements in the fox genome. Mamm Genome 3:233–236

Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A (2007) The delayed rise of present-day mammals. Nature 446:507–512

Bonaccorsi S, Gatti M, Pisano C, Lohe A (1990) Transcription of a satellite DNA on two Y chromosome loops of *Drosophila melanogaster*. Chromosoma 99:260–266

Bouzinba-Segard H, Guais A, Francastel C (2006) Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proc Natl Acad Sci USA 103:8709–8714

Bulazel K, Metcalfe C, Ferreri GC, Yu J, Eldridge MD, O'Neill RJ (2006) Cytogenetic and molecular evaluation of centromere-associated DNA sequences from a marsupial (Macropodidae: Macropus rufogriseus) X chromosome. Genetics 172:1129–1137

Bulazel KV, Ferreri GC, Eldridge MD, O' Neill RJ (2007) Species-specific shifts in centromere sequence composition are coincident with breakpoint reuse in karyotypically divergent lineages. Genome Biol 8:R170

Carone DM, Longo MS, Ferreri GC, Hall L, Harris M, Shook N, Bulazel KV, Carone BR, Obergfell C, O'Neill MJ, O'Neill RJ. (2009) A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. Chromosoma Feb; 118(1):113–25

Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell 14:1691–1704

Choo K (1997a) The centromere. Oxford University Press, Oxford

Choo KH (1997b) Centromere DNA dynamics: latent centromeres and neocentromere formation. Am J Hum Genet 61:1225–1233

Chueh AC, Wong LH, Wong N, Choo KH (2005) Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. Hum Mol Genet 14:85–93

Coffin JM, Hughes SH, Varmus HE (1997) Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Csink AK, Henikoff S (1998) Something from nothing: the evolution and utility of satellite repeats. Trends Genet 14:200–204

Dawe RK (2003) RNA interference, transposons, and the centromere. Plant Cell 15:297–301

Diaz MO, Barsacchi-Pilone G, Mahon KA, Gall JG (1981) Transcripts from both strands of a satellite DNA occur on lampbrush chromosome loops of the newt Notophthalmus. Cell 24:649–659

du Sart D, Cancilla MR, Earle E, Mao JI, Saffery R, Tainton KM, Kalitsis P, Martyn J, Barry AE, Choo KH (1997) A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. Nat Genet 16:144–153

Eichler EE (1999) Repetitive conundrums of centromere structure and function. Hum Mol Genet 8:151–155

Eldridge MD, Close RL (1993) Radiation of chromosome shuffles. Curr Opin Genet Dev 3:915–922

Elgin SC, Grewal SI (2003) Heterochromatin: silence is golden. Curr Biol 13:R895–R898

Epstein LM, Mahon KA, Gall JG (1986) Transcription of a satellite DNA in the newt. J Cell Biol 103:1137–1144

Ferreri GC, Marzelli M, Rens W, O'Neill RJ (2004) A centromere-specific retroviral element associated with breaks of synteny in macropodine marsupials. Cytogenet Genome Res 107:115–118

Ferreri GC, Liscinsky DM, Mack JA, Eldridge MD, O'Neill RJ (2005) Retention of latent centromeres in the Mammalian genome. J Hered 96:217–224

Fukagawa T, Nogami M, Yoshikawa M, Ikeno M, Okazaki T, Takami Y, Nakayama T, Oshimura M (2004) Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nature Cell Biol 6:784–791

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem 278:4035–4040

Garagna S, Zuccotti M, Capanna E, Redi CA (2002) High-resolution organization of mouse telomeric and pericentromeric DNA. Cytogenet Genome Res 96:125–129

Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI (2002) Establishment and maintenance of a heterochromatin domain. Science 297:2232–2237

Hammond SM, Bernstein E, Beach D, Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature 404:293–296

Hayman DL (1977) Chromosome number-constancy and variation. In Gilmore D (ed) The biology of Marsupials. Macmillan, London

Hayman DL (1990) Marsupial cytogenetics. In Cooper DW (ed) Mammals from pches and eggs: genetics, breeding and evolution of Marsupials and Monotremes. CSIRO, Melbourne

Heard E (2005) Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. Curr Opin Genet Dev 15:482–489

Henikoff S, Malik HS (2002) Selfish drivers. Nature 417:227

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098–1102

Hudson DF, Fowler KJ, Earle E, Saffery R, Kalitsis P, Trowell H, Hill J, Wreford NG, de Kretser DM, Cancilla MR, Howman E, Hii L, Cutts SM, Irvine DV, Choo KH (1998) Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. J Cell Biol 141:309–319

Jiang J, Birchler JA, Parrott WA, Dawe RK (2003) A molecular view of plant centromeres. Trends Plant Sci 8:570–575

Jin W, Melo JR, Nagaki K, Talbert PB, Henikoff S, Dawe RK, Jiang J (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. Plant Cell 16:571–581

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev 19:489–501

Kapitonov VV, Jurka J (1999) Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica 107:27–37

Kipling D, Warburton PE (1997) Centromeres, CENP-B and tigger too. Trends Genet 13:141–145

Knegt AC, Li S, Engelen JJ, Bijlsma EK, Warburton PE (2003) Prenatal diagnosis of a karyotypically normal pregnancy in a mother with a supernumerary neocentric 13q21 → 13q22 chromosome and balancing reciprocal deletion. Prenat Diagn 23:215–220

Kuznetsova I, Podgornaya O, Ferguson-Smith MA (2006) High-resolution organization of mouse centromeric and pericentromeric DNA. Cytogenet Genome Res 112:248–255

Lachner M, Jenuwein T (2002) The many faces of histone lysine methylation. Curr Opin Cell Biol 14:286–298

Lachner M, O'Sullivan RJ, Jenuwein T (2003) An epigenetic road map for histone lysine methylation. J Cell Sci 116:2117–2124

Lam AL, Boivin CD, Bonney CF, Rudd MK, Sullivan BA (2006) Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. Proc Natl Acad Sci USA 103:4186–4191

Lee HR, Neumann P, Macas J, Jiang J (2006) Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. Mol Biol Evol 23:2505–2520

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol 13:1192–1200

Li YC, Lee C, Hsu TH, Li SY, Lin CC (2000) Direct visualization of the genomic distribution and organization of two cervid centromeric satellite DNA families. Cytogenet Cell Genet 89:192–198

Li YX, Kirby ML (2003) Coordinated and conserved expression of alphoid repeat and alphoid repeat-tagged coding sequences. Dev Dyn 228:72–81

Lo AW, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E, Magliano DJ, Choo KH (2001a) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. EMBO J 20:2087–2096

Lo AW, Magliano DJ, Sibson MC, Kalitsis P, Craig JM, Choo KH (2001b) A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. Genome Res 11:448–457

Lowry PS, Eldridge MDB, Johnston PG (1994) Genetic analysis of a female macropodid hybrid (*macropus agilis* X *M. rufogriseus*) and her backcross offspring. Aus Mammal 18:79–82

Maison C, Bailly D, Peters AH, Quivy JP, Roche D, Taddei A, Lachner M, Jenuwein T, Almouzni G (2002) Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat Genet 30:329–334

Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics 157:1203–1208

Malik HS, Henikoff S (2002) Conflict begets complexity: the evolution of centromeres. Curr Opin Genet Dev 12:711–718

Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J Cell Biol 109:1963–1973

Mellone BG, Allshire RC (2003) Stretching it: putting the CEN(P-A) in centromere. Curr Opin Genet Dev 13:191–198

Metcalfe CJ, Bulazel KV, Ferreri GC, Schroeder-Reiter E, Wanner G, Rens W, Obergfell C, Eldridge MD, O'Neill RJ (2007) Genomic instability within centromeres of interspecific marsupial hybrids. Genetics 177:2507–2517

Miyahara M, Sumiyoshi H, Yamamoto M, Endo H (1985) Strand specific transcription of satellite DNA I in rat ascites hepatoma cells. Biochem Biophys Res Commun 130:897–903

Nagaki K, Song J, Stupar RM, Parokonny AS, Yuan Q, Ouyang S, Liu J, Hsiao J, Jones KM, Dawe RK, Buell CR, Jiang J (2003) Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. Genetics 163:759–770

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145

Nagaki K, Neumann P, Zhang D, Ouyang S, Buell CR, Cheng Z, Jiang J (2005) Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. Mol Biol Evol 22:845–855

Nakagawa H, Lee JK, Hurwitz J, Allshire RC, Nakayama J, Grewal SI, Tanaka K, Murakami Y (2002) Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. Genes Dev 16:1766–1778

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292:110–113

Neumann P, Yan H, Jiang J (2007) The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. Genetics 176:749–761

O'Neill RJ, O'Neill MJ, Graves JA (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid [see comments]. Nature 393:68–72

O'Neill RJ, Eldridge MD, Graves JA (2001) Chromosome heterozygosity and *de novo* chromosome rearrangements in interspecific mammalian hybrids. Mammalian Genome 12:256–259

O'Neill RJ, Eldridge MD, Metcalfe CJ (2004) Centromere dynamics and chromosome evolution in marsupials. J Hered 95:375–381

Ohzeki J, Nakano M, Okada T, Masumoto H (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J Cell Biol 159:765–775

Partridge JF, Borgstrom B, Allshire RC (2000) Distinct protein interaction domains and protein spreading in a complex centromere. Genes Dev 14:783–791

Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schöfer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, Jenuwein T (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell 107:323–337

Pidoux AL, Allshire RC (2004) Kinetochore and heterochromatin domains of the fission yeast centromere. Chromosome Res 12:521–534

Renault S, Rouleux-Bonnin F, Periquet G, Bigot Y (1999) Satellite DNA transcription in Diadromus pulchellus (Hymenoptera). Insect Biochem Mol Biol 29:103–111

Renfree MB (2006) Society for Reproductive Biology Founders' Lecture 2006 – life in the pouch: womb with a view. Reprod Fertil Dev 18:721–734

Rens W, O'Brien PC, Yang F, Graves JA, Ferguson-Smith MA (1999) Karyotype relationships between four distantly related marsupials revealed by reciprocal chromosome painting. Chromosome Res 7:461–474

Rens W, O'Brien PC, Fairclough H, Harman L, Graves JA, Ferguson-Smith MA (2003) Reversal and convergence in marsupial chromosome evolution. Cytogenet Genome Res 102:282–290

Rhoades MM, Dempsey E (1966) The effect of abnormal chromosome 10 on preferential segregation and crossing over in maize. Genetics 53:989–1026

Rice JC, Briggs SD, Ueberheide B, Barber CM, Shabanowitz J, Hunt DF, Shinkai Y, Allis CD (2003) Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. Mol Cell 12:1591–1598

Rofe RH (1979) G-banding and chromosomal evolution in Australian Marsupials. University of Adelaide, Adelaide

Rouleux-Bonnin F, Renault S, Bigot Y, Periquet G (1996) Transcription of four satellite DNA subfamilies in Diprion pini (Hymenoptera, Symphyta, Diprionidae). Eur J Biochem 238:752–759

Rudert F, Bronner S, Garnier JM, Dollé P (1995) Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mamm Genome 6:76–83

Sharman GB, Close RL, Maynes GM (1990) Chromosomal evolution, phylogeny and speciation of rock wallabies (*Petrogale*: Macropodidae). Aust J Zool 37:351–363

Singer MF (1982) Highly repeated sequences in mammalian genomes. Int Rev Cytol 76:67–112

Sullivan BA, Karpen GH (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat Struct Mol Biol 11:1076–1083

Sullivan BA, Willard HF (1998) Stable dicentric X chromosomes with two functional centromeres. Nat Genet 20:227–228

Sullivan KF, Hechenberger M, Masri K (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol 127:581–592

Svartman M, Vianna-Morgante AM (1999) Comparative genome analysis in American marsupials: chromosome banding and in-situ hybridization. Chromosome Res 7:267–275

Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an Arabadopsis histone H3 variant. Plant Cell 14:1053–1066

Topp CN, Zhong CX, Dawe RK (2004) Centromere-encoded RNAs are integral components of the maize kinetochore. Proc Natl Acad Sci USA 101:15986–15991

Ugarkovic D (2005) Functional elements residing within satellite DNAs. EMBO Rep 6:1035–1039

Valgardsdottir R, Chiodi I, Giordano M, Cobianchi F, Riva S, Biamonti G (2005) Structural and functional characterization of noncoding repetitive RNAs transcribed in stressed human cells. Mol Biol Cell 16:2597–2604

Ventura M, Mudge JM, Palumbo V, Burn S, Blennow E, Pierluigi M, Giorda R, Zuffardi O, Archidiacono N, Jackson MS, Rocchi M (2003) Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. Genome Res 13(9):2059–2068

Ventura M, Weigl S, Carbone L, Cardone MF, Misceo D, Teti M, D'Addabbo P, Wandall A, Björck E, de Jong PJ, She X, Eichler EE, Archidiacono N, Rocchi M (2004) Recurrent sites for new centromere seeding. Genome Res 14:1696–1703

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297:1833–1837

Volpe TA, Schramke V, Hamilton GL, White SA, Teng G, Martienssen RA, Allshire RC (2003) RNA interference is required for normal centromere function in fission yeast. Chromosome Res 11:137–146

Wang WaHL (2000) Rapid and parallel chromosomal number reductions in muntjac deer inferred from mitochondrial DNA phylogeny. Mol Biol Evol 17:1326–1333

Warburton PE (2004) Chromosomal dynamics of human neocentromere formation. Chromosome Res 12:617–626

Warburton PE, Cooke CA, Bourassa S, Vafa O, Sullivan BA, Stetten G, Gimelli G, Warburton D, Tyler-Smith C, Sullivan KF, Poirier GG, Earnshaw WC (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr Biol 7:901–904

White SA, Allshire RC (2004) Loss of Dicer fowls up centromeres. Nat Cell Biol 6:696–697

Willard HF (1990) Centromeres of mammalian chromosomes. Trends Genet 6(12):410–416

Williams BC, Murphy TD, Goldberg ML, Karpen GH (1998) Neocentromere activity of structurally acentric mini-chromosomes in Drosophila. Nat Genet 18:30–37

Wong LH, Choo KH (2004) Evolutionary dynamics of transposable elements at the centromere. Trends Genet 20:611–616

Wu ZG, Murphy C, Gall JG (1986) A transcribed satellite DNA from the bullfrog Rana catesbeiana. Chromosoma 93:291–297

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335–340

Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y, Wing RA, Han B (2004) Structural features of the rice chromosome 4 centromere. Nucl Acids Res 32:2023–2030

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

Chapter 5
Evolutionary New Centromeres in Primates

Mariano Rocchi, Roscoe Stanyon, and Nicoletta Archidiacono

Contents

Abstract The centromere has a pivotal role in structuring chromosomal architecture, but remains a poorly understood and seemingly paradoxical "black hole." Centromeres are a very rapidly evolving segment of the genome and it is now known that centromere shifts in evolution are not rare and must be considered on a par with other chromosome rearrangements. Recently, unprecedented findings on neocentromeres and evolutionary new centromeres (ENC) have helped clarify the relationship of the

M. Rocchi (✉) and N. Archidiacono
Dipartimento di Genetica e Microbiologia, Via Amendola 165/A, 70126 Bari, Italy
e-mail: rocchi@biologia.uniba.it

R. Stanyon
Dipartimento di Biologia Evoluzionistica, Via del Proconsolo 12, 50122 Firenze, Italy

Đ. Ugarković (ed.), *Centromere,* Progress in Molecular and Subcellular Biology 48,
DOI: 10.1007/978-3-642-00182-6_5, © Springer-Verlag Berlin Heidelbarg 2009

centromere within the genome and shown that these two phenomena are two faces of the same coin. No prominent sequence features are known that promote centromere formation and both types of new centromeres are formed epigenetically, both clinical neocentromeres and ENC cluster at chromosomal "hotspots." The clustering of neocentromeres in 8p is probably the result of the relatively high frequency of non-canonical pairing. Studies on the evolution of the chromosomes 3, 13, and 15 help explain why there are clusters of neocentromeres. These domains often correspond to ancestral inactivated centromeres and some regions can preserve features that trigger neocentromere emergence over tens of millions of years. Neocentromeres may be correlated with the distribution of segmental duplications (SDs) in regions of extreme plasticity that often can be characterized as gene deserts. Further, because centromeres and associated pericentric regions are dynamically complex, centromere shifts may turbocharge genome reorganization by influencing the distribution of heterochromatin. The "reuse" of regions as centromere seeding-points in evolution and in human clinical cases further extends the concept of "reuse" of specific domains for "chromosomal events."

5.1 The "Black Hole"

The centromere, a term coined by Darlington 1936, is the primary constriction where the kinetochore forms and the spindle fiber attaches to ensure correct chromatid segregation during cell division. The centromere has always been given a pivotal role in structuring chromosomal architecture, and classical analyses emphasized Robertsonian fissions and fusions as well as pericentric inversions as the principle mechanisms in the transformation of species diploid (2n) and fundamental numbers (FN, number of chromosome arms). More recent investigations have also paid attention to deletions, duplications, tandem fusions, and centromere shifts, with both the deactivation and the activation of centromere playing a fundamental role. The pericentromeric regions of centromeres are regions rich in duplicons, transposons, retro elements, and even pseudogenes and expressed genes. They are hot spots of chromosome changes in both evolution and in disease (Villasante et al. 2007).

Clearly then, the centromere is a key structure in the evolution of eukaryotic chromosomes, yet remains poorly understood and seemingly paradoxical. Early work suggested that particular satellite sequences were involved in centromere formation but the comparative study of centromere DNA showed that it was highly variable across species (O'Neill et al. 2004) (see Chaps. 2–4 of this book).

In the last years, unprecedented findings on neocentromeres and evolutionary new centromeres (ENC) added additional oddities to this "black hole" of biology. On the other hand, they started to clarify the complex relationship of the centromere with the underlying sequences. Montefalcone et al. (1999) showed that a

centromere, during evolution, can move along the chromosome without any accompanying chromosomal rearrangements. This unusual centromere behavior is now well documented in a large array of taxa, in particular, primates. It was also shown that ENCs have an intriguing connection with a related phenomenon: human clinical neocentromeres. This chapter mainly addresses the evolutionary aspects of neocentromeres, but ENCs and neocentromeres are, very likely, two faces of the same phenomenon. For this reason, the clinical neocentromeres will be briefly summarized in the following paragraph. For an exhaustive review see Marshal et al. (2008)

5.1.1 Human Clinical Neocentromeres

Neocentromeres are analphoid centromeres that emerge in ectopic chromosomal regions. The emergence of a neocentromere most frequently occurs to provide mitotic stability to otherwise acentric chromosome fragments resulting from a rearrangement (Amor and Choo 2002; Warburton 2004; Marshall et al. 2008). The stabilized supernumerary chromosome has detrimental phenotypic consequences, and it is usually discovered when these clinical patients are examined cytogenetically.

Nearly 100 such cases were reported in the literature (cf. Marshall et al. 2008). Marshall et al. (2008) report that clinical neocentromere are noted once in every 70,000–200,000 live births, but these studies do not include the incidence of balanced rearrangements which have no phenotypic consequences and are not caught by the clinical filter (see Capozzi et al. 2008). Sometimes balanced neocentromeres are serendipitously found in normal individuals (see below). The chromosomal distribution of neocentromeres is reported in Fig. 5.1.

As mentioned, neocentromere emergence is usually an opportunistic, secondary event, concomitant to a rearrangement that generated an acentric fragment. This implies that human clinical neocentromeres are not the consequence of any kind of sequence transposition or mutational modification, and that, consequently, these events are epigenetic in nature (Alonso et al. 2003) (see also Chap. 1 of this book).

The chromosomal localization of neocentromeres (see Fig. 5.1) has usually been attained by fluorescence in situ hybridization (FISH) using BAC or similar DNA probes, with the aim of identifying clones mapped to opposite sides of the centromere. Occasionally, this approach for various reasons provided only an approximate mapping. One reason is that the neocentromere does not contain a heterochromatic block that can be very helpful in orienting the probe hybridization to one side or the other of the centromere. Additionally, several supernumerary, neocentromeric chromosomes have an inverted-duplication structure that makes characterization difficult. Neocentromeres in small ring chromosomes are also difficult to map because the primary constriction is not easily identified. These limitations explain why the mapping of chromosomal regions harboring neocentromeres was sometimes fairly approximate (see Fig. 5.1). In some instances, however, the neocentromere was

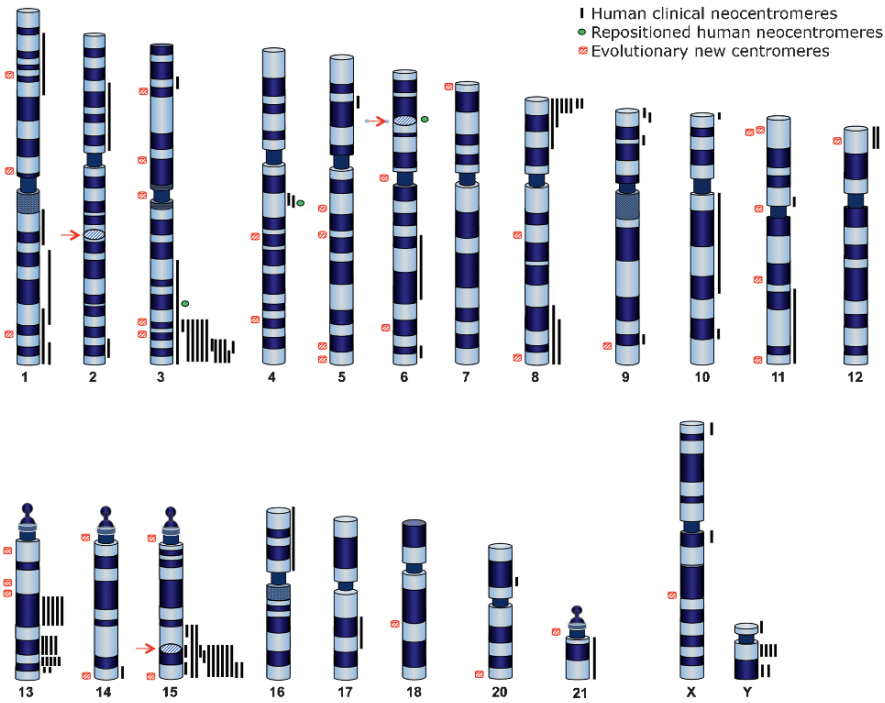


Fig. 5.1 ENCs and neocentromeres. The ideograms graphically report human clinical centromeres, represented by a *black bar* spanning the seeding point, on the right of each chromosome (modified from Marshall et al. 2008). The figure includes 1 new clinical centromere reported on chromosome 9 (Capozzi et al. 2008) and one repositioned centromere on chromosome 6 (Capozzi et al. in press). The localization appears very approximate in some instances, for the reasons discussed in the text (Sect. 5.1.1). The three repositioned centromeres found in normal persons are represented by a small *green circle*. ENCs are indicated, *in red*, on the left of the chromosomes. *Red arrows* indicate inactivated ancestral centromeres. The Supplement Table 5.1 (see at the end of this chapter) reports in detail the data graphically summarized in this figure

mapped down to the sequence level using a ChIP-on-chip approach. In this method, living cells are crosslinked in situ by adding formaldehyde. DNA is then sheared by sonication and immunoprecipitated using antibodies against centromeric proteins (CENP, usually CENP-A and CENP-C). Purified DNA fragments are then amplified, labeled, and hybridized to a high density BAC or oligo arrays (see Capozzi et al. 2008). Thirteen neocentromeres were precisely mapped in this way (Lo et al. 2001a, b; Alonso et al. 2003, 2007; Saffery et al. 2003; Sumer et al. 2003; Chueh et al. 2005; Cardone et al. 2006; Capozzi et al. 2008). The CENP domain ranged from ~54 to 450 kb. The size can be occasionally over-estimated if BAC arrays are used. Sequence comparison among these regions did not show any prominent features that could be predictive of centromere-forming potential. In other words, it is not

evident what makes a sequence "centromere competent." Another complication is the striking difference between a "normal" centromere, up to 3–4 Mb in size, and neocentromeres composed of as low as 50 kb of "plain" sequence. It has to be noted, however, that the frequently reported mosaicism suggests that neocentromeres are not so efficient. This point will be further discussed below.

The phenotypic problems inherent in patients with neocentromeres also imply that they have no evolutionary future. It can be easily hypothesized that the fitness of these individuals is negligible. The neocentromere-ENC connection could therefore appear problematic. However, some recent lines of evidence suggested a surprisingly strong relationship. For instance, same chromosomal domain can be used as seeding-point for both neocentromeres and ENCs. A second line of evidence revealed that some seeding-point domains correspond to ancestrally inactivated centromeres (see below). Lastly, three familial cases of human neocentromeres were discovered segregating in perfectly normal people (Amor et al. 2004; Ventura et al. 2004; Capozzi et al. in press). These three cases can be considered as repositioned centromeres "in progress." They are familiarly inherited and have no phenotypic implications; indeed their discovery was accidental.

5.2 Evolutionary Repositioned Centromeres in Primates

Karyotype evolution has been mainly studied using whole-chromosome painting probes. This approach has the advantage of mapping translocation differences between species, but does not usually provide information on intrachromosomal rearrangements or marker order differences. Recently, the availability of large cloned DNA collections of BACs and fosmids (see P. de Jong lab at http://bacpac.chori.org/home.htm; see also paragraph 9.6, Technical note) made it possible to study by FISH marker order changes during evolution in chromosomes of different species (molecular cytogenetic approach). The precise mapping of thousands of clones is graphically displayed in genome browsers (see the track "BAC End Pairs" or "Fosmid End Pairs" in UCSC, for instance). Two or more BAC clones can be simultaneously hybridized and their reciprocal order can be unequivocally defined. This cytogenetic approach to synteny definition complements other approaches that have been exploited to define genome organization: radiation hybrid mapping, linkage analysis, and sequencing (see Rocchi et al. 2006). Importantly, the molecular cytogenetic approach is sequence independent, and it can substantially aid sequence assembly, because the pure shot-gun approach, used for most genomes, is error prone (Green 1997; Roberto et al. 2008). For a fine synteny definition of complex genomes using the molecular cytogenetics technology, see Roberto et al. (2007) and Misceo et al. (2008) and the corresponding Web pages http://www.biologia.uniba.it/lar/ and http://www.biologia.uniba.it/gibbon/, respectively, provided as Supplemental Material to these publications.

Synteny arrangement comparisons allowed Montefalcone et al. (1999), as mentioned earlier, to disclose that some centromeres shifted along the chromosome during evolution. Studies over the last decade have amply demonstrated that centromere shifts in evolution are not rare and must be considered on a par with other chromosome rearrangements such as translocations, inversion, duplications, and deletions. Ventura et al. (2007), comparing human and macaque, clarified how very frequent ENC are in primate evolution. In total, between macaque and humans there are 14 ENC; nine ENCs occurred in macaque lineage and five occurred in the human lineage. The last common ancestor of macaques and humans is estimated at about 25 million years ago (mya). So ENC in this case formed about once every three million years. Perhaps surprisingly, by comparison in the same arch of time, there are only four translocation differences (about one translocation every 12 million years). We might conclude from this example that ENC are four times more frequent than cytogenetically visible translocations and represent a significant facet of mammalian chromosomal evolution. ENCs were reported in the evolution of chromosome 3 (Ventura et al. 2004), chromosome 6 (Eder et al. 2003), chromosome 10 (Carbone et al. 2002), chromosome 11 (Cardone et al. 2007), chromosome 13 (Cardone et al. 2006), chromosome 14 and 15 (Ventura et al. 2003), chromosome 20 (Misceo et al. 2005), and chromosome X (Ventura et al. 2001). Figure 5.1 graphically reports, on the left of each chromosome, all the published ENCs. Supplement Table 5.1 (see at the end of this chapter) reports details of neocentromeres and ENCs literature data. It is interesting to note that the centromere is apparently a very rapidly evolving segment of the genome. Further, because centromeres and associated pericentric regions are dynamically complex, centromere shifts may turbocharge genome reorganization by influencing the distribution of heterochromatin (Ishii et al. 2008).

5.3 Hotspots of Neocentromere Formation

A clearly recognizable trend from the human clinical cytogenetic data is the clustering of neocentromere formation sites at chromosomal "hotspots." Certain regions of chromosomes – for example, 3q, 8p, 13q, and 15q telomeric regions – seem particularly prone to forming neocentromeres (Fig. 5.1). The survival of individuals with more distal inverted duplications will be favored (as such individuals possess a smaller region of partial trisomy or tetrasomy); it is therefore logical that neocentromeres cluster around the distal ends of chromosomes. It follows that some other regions with neocentromere-forming potentiality have never been described because of this bias. What becomes fixed in evolution is, therefore, the end result of mutation and the selectional filter. The neocentromere reported at 9q33.1 is paradigmatic in this respect (Capozzi et al. 2008). The propositus, in fact, was found to carry an interstitial deletion of chromosome 9,

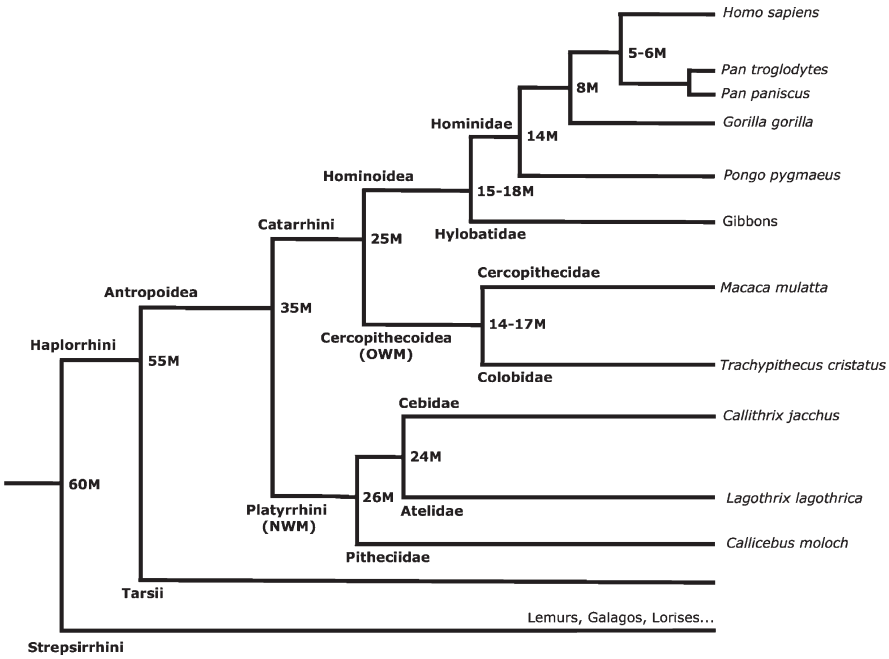


Fig. 5.2 Phylogeny of primates. Summary of the phylogenetic relationship among extant primates. Branching time is according to Raaum et al. (2005) and Opazo et al. (2006). The bars' length is not proportional to elapsed time. The figures indicate the branching time in million years

of about 12 Mb (9q31.3-9q33.1). The parents were investigated because of the deletion in the son. The mother had a small ring chromosome that resulted from the excision of the 12 Mb from the chromosome 9. A neocentromere at 9q33.1 had stabilized the ring chromosome. The son had inherited the deleted chromosome but not the ring. This neocentromere would have been never detected if malsegregation had not occurred. No such neocentromere was detected in supernumerary chromosomes.

Studies on the evolution of the chromosomes where clustering of neocentromeres were reported (3q, 13q, and 15q) put these regions in a completely new light. These chromosomes were investigated in detail, and each of these clusters disclosed distinct, intriguing aspects of the relationship between human clinical neocentromeres and ENCs. For this reason they will be described in detail later.

The full appreciation of these data presupposes a basic knowledge of primate phylogeny, which is summarized in Fig. 5.2. It is also important that the reader is acquainted with the concept of the "outgroup" in phylogenetic studies. A brief description is reported in the Sect. 5.6.

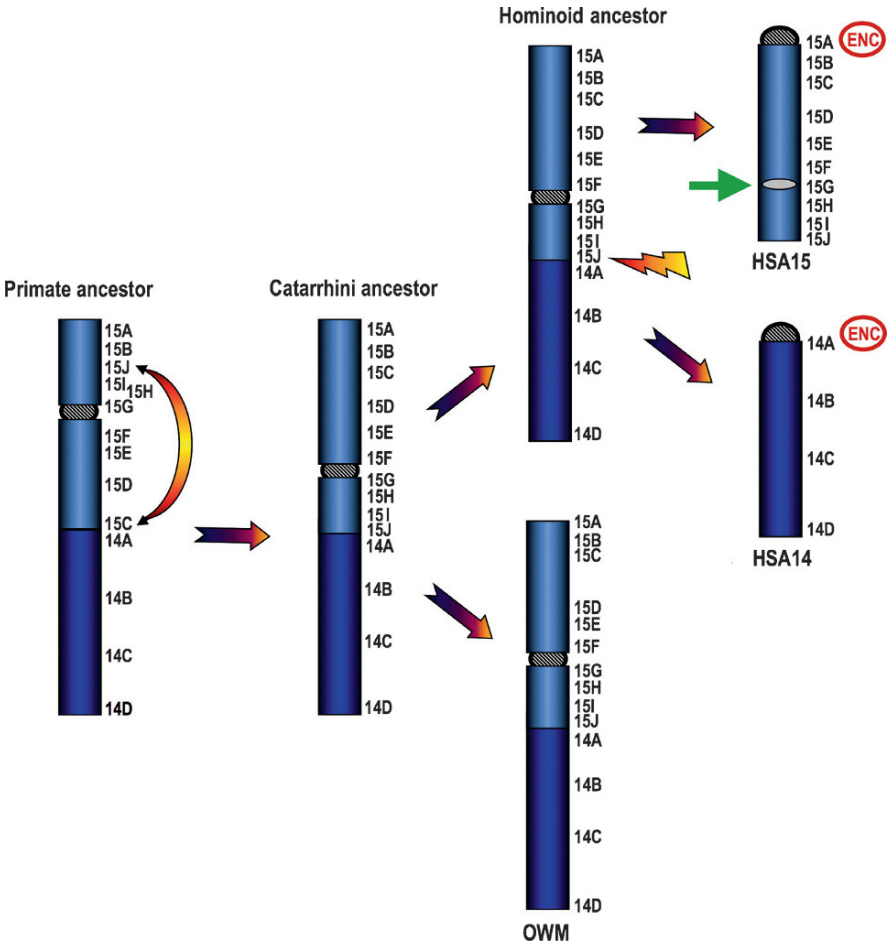


Fig. 5.3 Evolution of human chromosomes 15 and 14. The figure delineates the evolutionary history of chromosomes 15 and 14 in OWMs and Hominoidea. BAC clones used in the synteny investigation are represented by letters on the right of the chromosomes. The letter-BAC correspondence is reported in Supplement Table 5.1. Chromosomes 15 and 14 in Hominoidea were generated by fission of an ancestral chromosome, which appears to be composed of these two chromosomes arranged head-tail. ENC in a *red circle* indicates the emergence of an evolutionary new centromere. The *green arrow* points to the inactivated centromere. For details see text

5.3.1 Evolution of Chromosome 15

Human chromosomes 15 and 14 derive from the fission of an ancestral chromosome in the Hominoidea ancestor. Comparison with outgroup species confirms that the fission is the derivative rearrangment. Figure 5.3 reports the study of the evolution of these chromosomes through the use of BAC clones that showed that the marker

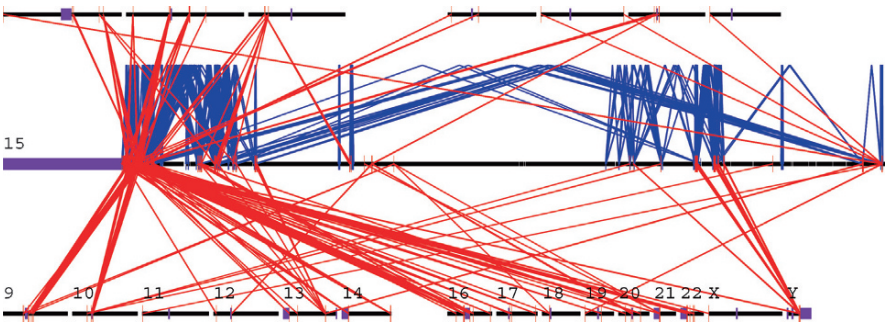


Fig. 5.4 Segmental duplication analysis of chromosome 15. The figure illustrates the interchromosomal (*red lines*) and intrachromosomal (*blue lines*) segmental duplications of chromosome 15 (Courtesy of Dr. E.E. Eichler; from Bailey et al. 2002)

order was perfectly conserved between macaque chromosome 7 (*Macaca mulatta*, MMU) and the two human chromosomes. To derive the two independent human chromosomes, 14 and 15, you only need to fission between markers F and G (Ventura et al. 2003) (Fig. 5.3). One novel centromere emerged in human chromosome 15, corresponding to the telomeric region of the short arm of MMU7 (Fig. 5.3). A second centromere emerged on chromosome 14 and corresponded to the fission point of MMU7. The ancestral centromere, precisely mapped by the apparent split of marker E (chr15:82,835,478-83,006,963, UCSC, March 2006 release), got inactivated.

Segmental duplications (SDs) are biased against pericentromeric regions (She et al. 2004). The graphic representation of the distribution of SDs of chromosome 15 shows a clear clustering of SDs at 15q24-26 (Fig. 5.4). In light of the evolutionary analysis of chromosome 15 we have reported, they represent the remains of the pericentromeric SDs that flanked the ancestral centromere. No alphoid sequences are present in this domain, suggesting that the loss of this satellite DNA, typical of primate centromeres, was relatively rapid. The most interesting observation, however, is that human clinical neocentromeres clustering at 15q24-26 perfectly overlap the distribution of SDs. Apparently, the region has preserved features that trigger neocentromere emergence. This potentiality has been conserved for approximately 25 MY, the time of divergence between Hominoidea from Cercopithecoidea (Old World Monkeys, OWM) (Raaum et al. 2005).

Main conclusions are as follows: (i) neocentromeres can emerge in domains corresponding to ancestral inactivated centromeres; (ii) neocentromeres are scattered over a fairly relatively large area (15q24-26), overlapping the dispersion of SDs; (iii) apparently, centromere forming latency is not linked to a specific sequence.

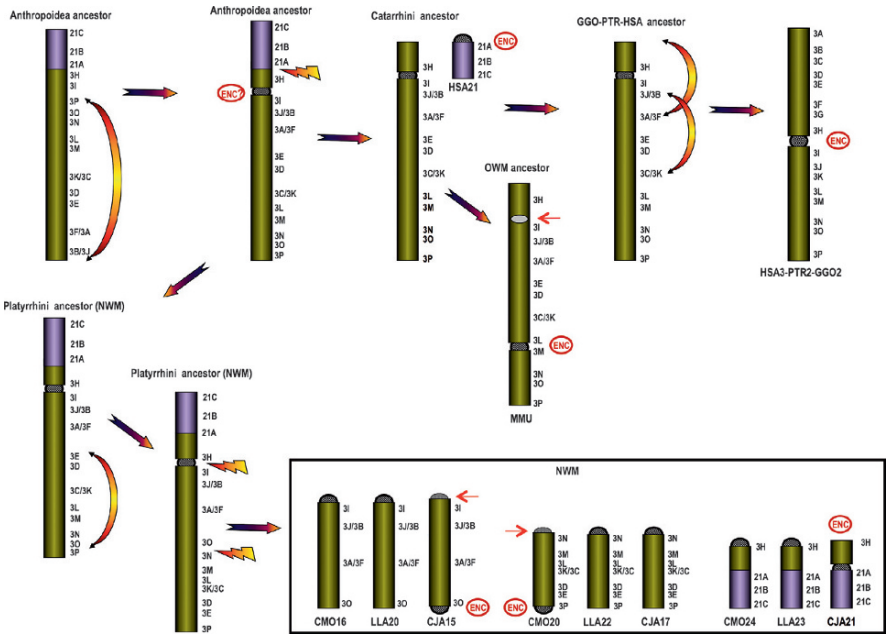


Fig. 5.5 Evolution of chromosome 3. Delineation of the chromosomal changes of chromosome 3 during primate evolution, modified from Ventura et al. (2004). BAC clones used in the synteny investigation are represented by letters on the right of the chromosomes. The letter-BAC correspondence is reported in Supplement

5.3.2 Evolution of Chromosome 3

The evolutionary history of chromosome 3 is relatively complex in comparison to that of 15/14 (Ventura et al. 2004). Figure 5.5 shows how the human chromosome 3 can be derived from the primate ancestor by fission of the 21 synteny and several inversions. Marker order comparison among selected primate species revealed that the centromeres in both Hominoidea and OWM are ENCs. The paucity of SDs around this ENC (She et al. 2004) could be interpreted as the consequence of its recent origin. We had the opportunity to study one case of a neocentromere that resulted from the excision of a small region, including the centromere, to form a small autonomous chromosome (Wandall et al. 1998). The neocentromere appeared located in a domain almost overlapping with the ENC described in macaque (Ventura et al. 2004).

Main conclusion: the same chromosomal domain was used as a seeding point for an ENC and for a human clinical neocentromere.

5.3.3 Evolution of Chromosome 13

Contrary to chromosome 3, chromosome 13 can be regarded as one of the most evolutionary conserved chromosomes. The human form very likely corresponds to that of the primate ancestor, which in turn differs from the mammalian ancestor form just for a small inversion (Cardone et al. 2006). The same syntenic arrangement of the mammalian ancestor was found in chicken (Consortium 2004) that diverged from mammals about 310 mya. In OWMs, a novel centromere emerged in a region in the middle of the long arm (13q21). Interestingly, a similarly located, independent ENC emergence was detected in pigs. Additionally, some human neocentromeres reported on chromosome 13 mapped close to the same chromosomal domain. These findings resemble the results reported for chromosome 3. The study, however, exposed some important additional aspects of the centromere repositioning phenomenon: (i) this region maintained centromere forming potential for a very long time of about 95 my, that is, the divergence time of Cetartiodactyla and Primates; (ii) human probes mapping in the seeding region had a very variable results on different OWM species (MMU, *Papio hamadryas*, *Trachypithecus cristatus*, and *Chlorocebus aethiops*), indicating that the region is extremely plastic; (iii) the ENC was seeded in a very large gene-desert region (4.88 Mb) (Lomiento et al. 2008). This last feature will be discussed in detail later.

5.3.4 Neocentromere Clustering at 8p

Contrary to chromosomes 15, 3, and 13, the evolutionary history of chromosome 8 did not reveal any feature that could be of help in interpreting the clustering of clinical neocentromeres at 8p (personal unpublished data). Recent studies published by Dr. Zuffardi's group on cytogenetic anomalies of 8p can be helpful to interpret this clustering. They found that parents of patients carrying de novo 8p chromosomal rearrangement, usually the mother, were heterozygous for an 8p23.1 inversion, delimited by two large clusters of olfactory receptor genes (Giglio et al. 2001). The noncanonical meiotic pairing, consisting in the refolding of one chromosome onto itself, favors the formation of derivative 8p chromosomes, including inv dup(8p) (see Fig. 5.4 of Giglio et al. 2001). The inversion is relatively common: 26% of the studied population appears heterozygous for the inversion and the neocentromere reports in literature are all acentric inv dup(8p) rescued by a neocentromere which insured their mitotic survival. Main conclusion: the reason for the clustering of neocentromeres in 8p is probably the result of the relatively high frequency of noncanonical pairing in individuals heterozygous for the 8p inversion.

An alternative hypothesis, discussed below, is that the potential restructuring of chromatin at the break that generated the inv dup(8) could be a concurrent epigenetic cause of neocentromere emergence.

5.3.5 *Reuse of Sites of "Chromosomal Events" in Evolution*

It is well known that the mouse genome accumulated a large number of chromosomal rearrangements during evolution (Waterston et al. 2002). Subsequent independent bioinformatic studies have shown, in humans, an extensive "reuse" of breakpoints (Pevzner and Tesler 2003; Murphy et al. 2005), and, additionally, an enrichment of segmental duplications in regions of synteny breaks between the human and mouse genomes (Armengol et al. 2003; Bailey et al. 2004). The SD in humans, however, occurred in the lineage leading to humans long after rodent/primates divergence. The conclusion was that the analysis "supports a nonrandom model of chromosomal evolution that implicates specific regions within the mammalian genome as having been predisposed to both recurrent small-scale duplication and large scale evolutionary rearrangements." The "reuse" of regions as centromere seeding-points in evolution and in human clinical cases further extends the concept of "reuse" of specific domains for "chromosomal events."

5.4 Human Repositioned Centromeres "in Progress"

A crossover inside the region encompassed by the normal and the repositioned centromere results in the formation of dicentric or acentric fragments. In contrast with the expectation that heterozygous carriers of neocentromeres have diminished fitness, the number of repositioned centromeres is relatively high and many repositioned centromeres have been fixed in different species. Meiotic drive in females, as reported for Robertsonian fusions in humans, in favor of the repositioned chromosome might be a possible explanation (Pardo-Manuel de Villena and Sapienza 2001). Meiotic drive has also been invoked to account for the progressive acquisition of heterochromatin in the neocentromeric regions (Henikoff et al. 2001). The progression towards normal centromere complexity, composed of large satellite DNA arrays, is presumed to stabilize neocentromere function. Most clinical neocentromeres are relatively unstable, as suggested by the fact that they are often found as mosaics. Population structure and genetic drift can also be hypothesized to have played an important role in neocentromere fixation.

It can be reasonably supposed, furthermore, that repositioned centromeres that reach fixation are only a minority of those that have emerged in the population. Repositioned centromeres have no clinical consequences. They therefore escape, in humans, the clinical filter that intercepts most of the neocentromeres present as supernumerary chromosomes. Prenatal cytogenetic analyses are most often performed without parental clinical indication. Further, centromere repositioning events can easily be misinterpreted as pericentric inversions. In non-human species, no cytogenetic population data are available, but the number of neocentromere that become fixed ENC is surely a minority. As a consequence, the number of centromere repositioning

events in both clinical and evolutionary cytogenetics must be much higher than that noted in the literature.

Examples of balanced centromere repositioning events with no obvious phenotypic effect do exist. The first instances were reported on the Y chromosome (Bukvic et al. 1996; Rivera et al. 1996; Tyler-Smith et al. 1999). The large block of heterochromatin present in this chromosome, however, hampered a full characterization of these repositioned centromeres, in which the satellite DNA could have played a nonminor role. More recently, three autosomal examples of repositioned centromeres have been reported at 3q24 (Ventura et al. 2004), 4q21.3 (Amor et al. 2004), and 6p22.1 (Capozzi et al. in press). They were serendipitously found (two because of a prenatal diagnosis). We will focus on the last case because it showed unprecedented features.

5.4.1 Repositioned Centromere at 6p22.1

The variant chromosome was discovered during a prenatal diagnosis Capozzi et al. (in press). Molecular cytogenetic analysis showed that the centromere was located in the middle of the short arm, at 6p22.1, without marker order changes. The analysis was extended to the family. The repositioned centromere was found in six individual in three generations. The segregation in three generation and the absence of any phenotypic problem suggested that the repositioned centromere was perfectly functional. In some metaphases, however, extra copies of chromosome 6 indicated that the functionality was not identical to a normal centromere. The precise position of the neocentromere was investigated using ChIP-on-chip analysis that indicated that it was located at chr6:26,407–26,491 kb. The evolutionary history of chromosome 6 had been already delineated by Eder et al. (2003), but the position of the centromere in the ancestor of primates could not be defined with certainty. New data accumulated in the literature allowed us to establish that the ancestral form of chromosome 6 in primates had the same marker order as in humans, but the centromere was located at 6p22.1. This centromere repositioned to the present-day location in the Hominoidea ancestor before gibbon branching, that is at least 17 mya (Raaum et al. 2005). The repositioned centromere was found about 2 Mb apart from the ancestral centromere. In our family case, therefore, it appears as if the centromere jumped back to the ancestral position, where it was located about 17 mya.

5.5 Evolutionary Fate of Novel Centromeres

The organization of a “mature” centromere is complex. In primates, the central core is composed of a large array of alpha satellite DNA, usually surrounded by a cluster of SDs. Occasionally, other types of satellite DNA flank the alphoid core.

Similarities with human clinical neocentromeres and human "repositioned" centromeres (see above) strongly suggest that the seeding event is epigenetic in nature, not accompanied by any sequence changes. In macaque, nine of 22 chromosomes are ENCs (Ventura et al. 2007). This subset of centromeres, however, is indistinguishable from the "normal" ones: all autosomal macaque centromeres possess a large block of alphoid DNA (Ventura et al. 2007). The same applies to the humans ENCs (Ventura et al. 2007). It appears as if the progression of these centromeres, from a "plain" sequence, obligatory ends in the acquisition of complexity. To better understand this process, it is worth noting that, as already mentioned, many human clinical neocentromeres and repositioned centromeres have been found to be mitotically unstable, with mosaicism, especially in supernumerary chromosomes (Marshall et al. 2008). Altogether, these observations suggest that rapid progression stabilizes the functionality of the centromere.

Data on pericentromeric SDs of repositioned centromeres are contrasting. Human centromeres of chromosome 3, 6, 11, 14, 15, and 21 are evolutionary new. While acrocentric chromosomes 14, 15, and 21 show large clusters of pericentromeric SDs, the centromere of chromosome 3 and 6 are relatively poor in SDs. Data on non-human primates are scarce, specifically because the shot-gun sequence approach is inefficient to spot SDs, especially if they are duplicated in tandem (Eichler 2001). Their characterization requires meticulous assembly efforts because of the homology, occasionally very high, of SDs. Using a combination of BAC library screening, FISH experiments, and STS sequencing, we were able to characterize the pericentromeric region of macaque ENC of chromosome 6. It appeared as if a 250-kb segment was imperfectly duplicated seven times around the macaque centromere (Ventura et al. 2007). Several deletions were supposed to have occurred during the process, because STSs failed several time to amplify the DNA of some macaque BACs.

Studies on the expression of genes embedded in human neocentromeres have shown that they are not affected by their unusual position (Wong and Choo 2001; Saffery et al. 2003; Capozzi et al., in press). However, the deep restructuring that accompanies neocentromere progression, as deduced from the results on MMU6 ENC, can be supposed to physically disrupt the sequence integrity of these genes and that a purifying selection would negatively affect the fixation in the population of these ENC. We tested this hypothesis by checking the gene density in the regions where ENC were seeded (Lomiento et al. 2008). The regions of ENCs seeding were significantly depleted of genes. It can be concluded that this circumstance had played a crucial role in their fixation in the population.

Further, we examined the occurrence of SDs around the ENCs present in humans and OWM. SDs in human have been characterized in great detail (She et al. 2004), but the macaque assembly is relatively poor in this respect. Using appropriate macaque BAC clones, we investigated SDs located pericentromerically to macaque ENCs. We found that all the examined regions have a certain level of SDs, but, as in humans, the amount varied considerably. The differences could not be attributed, in macaque, to the tempo of their seeding. All of them have been

seeded in the common ancestor of OWM, between 16 and 25 mya (Raaum et al. 2005). It could be hypothesized that the amount of SDs proceeds as a cascade process. In this case, pericentromeric regions with a higher amount of SDs should contain older SDs. To test this hypothesis would require, however, a substantial effort in sequencing these complex regions.

An additional interesting point of discussion is provided by the unusual findings reported on the pericentromeric region of macaque chromosome 13 (Cardone et al. 2006). The comparison of the different duplication pattern in three OWM species (*Macaca mulatta*, MMU, Cercopithecinae), sacred baboon (*Papio hamadryas*, PHA, Cercopithecinae), and silvered-leaf monkey (*Trachypithecus cristatus*, TCR, Colobinae) showed an unprecedented plasticity. The involved region spans about 3.7 Mb (from marker H2 to marker H8 in Fig. 5.2b of Cardone et al. (2006)). Importantly, this ENC was seeded in a vast gene desert as reported by Lomiento et al. (2008), and appears to involve almost the entire gene-desert, that is about 4.88 Mb. It could be hypothesized that the size of the gene desert defines the degree of plasticity of the pericentromeric region.

5.5.1 Telomeres, Centromeres, and Breakpoint Regions

Evolutionary studies of karyotypes have shown that chromosomes frequently result from the fission of ancestral chromosomes. In humans, chromosomes 15 and 14 and chromosome 21 among others were generated in this way (see above). In such instances at least one new centromere emerged at one telomere or at the breakpoint of the fission. One hypothesis on the origin of centromeres in eukaryotes is that they derived from telomeres. According to this hypothesis, telomeres existed before centromeres and that the recurrent appearance of unstable dicentric chromosomes through the formation of new centromeres (from telomeres) may have had a role in the origin of multiple chromosomes (Villasante et al. 2007). The evolution of chromosome 3 in NWM shows several examples of the centromere-telomere functional interchange that may be a remnant of the evolutionary origin of centromeres. The studied species were wooly monkey (*Lagothrix lagothricha*, LLA), common marmoset (*Callithrix jacchus*, CJA), dusky titi (*moloch*, CMO). The three segments of chromosome 3 in these NWM species had a similar marker content and orientation, but the centromere position was puzzling (Fig. 5.5b). The orthologous chromosomes LLA20 and CMO16 had the centromere telomerically located, close to marker I, while CJA15 centromere mapped at the opposite telomere, close to marker O. Similarly, the centromeres of CJA17 and LLA22 were located at one telomere, close to marker N, while in CMO the centromere was located at the opposite telomere, close to 3P. The three chromosomes were generated by two successive fissions. The first one occurred at the ancestral centromere, while the second mapped between the markers O and N. It is worth noting that both ends generated by the second

fission accommodated a centromere, and that the novel centromere in CJA21 appears to be located at the breakpoint region that, in Hominoidea, generated the human chromosome 21.

The two human clinical neocentromeres reported by Ventura et al. (2003) are invdup(15). It was hypothesized that breaks, through chromatin reorganization, could favor the emergence of neocentromeres. Literature data on breaks that generated the acentric fragments and neocentromere seeding-points, however, are relatively approximate. Precise mapping at the sequence level is mandatory to clarify this question. In the case of a neocentromere that stabilized the ring chromosome excised from chromosome 9, both the neocentromere and the breaks were precisely mapped (Capozzi et al. 2008; see above). They turned out to be about 2.1 Mb apart, which is in the range of the neocentromere-ENC correspondence reported so far.

5.5.2 ENCs in Non-Primate Mammals and in Other Taxa

The ENC phenomenon appears widespread in a large number of different taxa. In addition to primates, clear examples of ENCs are available for cattle (Larkin et al. 2003; Everts-van der Wind et al. 2005), pig (Cardone et al. 2006), rat (Kobayashi et al. 2008), birds (Kasai et al. 2003), and rice (Nagaki et al. 2004). For marsupials, see Chap. 4. One of the most interesting species, in this context, is the donkey. Comparison of donkey and zebra, using the horse as outgroup, revealed that at least five ENCs emerged in donkey (Carbone et al. 2006) but, because we were able to analyze only larger chromosomes for which marker order could be unequivocally established, there may be additional ENCs. These data are impressive if one considers that donkey and zebra diverged less than 1 mya (Oakenfull and Clegg 1998; Oakenfull et al. 2000).

5.5.3 Concluding Remarks

Centromeres, the "black hole" of the genome, even in the sequencing era resist easy explanation. Yet over the last decade, notable progress has been made especially using molecular cytogenetics. It has become increasingly clear that neocentromere formation and ENCs must be considered as important modes of genome evolution. Perhaps even more remarkable is that the mechanisms in the formation of both types of centromere are intimately related. The "reuse" of regions as centromere seeding-points in evolution and in human clinical cases further extends the concept of "reuse" of specific domains for "chromosomal events." Centromere-forming domains often correspond to ancestral inactivated centromeres and some regions

can preserve features that trigger neocentromere emergence over tens of millions of years of evolutionary time. In 2009, we will celebrate the 200th birthday of Charles Darwin and 150 years since the publication of his monumentus book "On the Origin of Species." We now can appreciate that centromeres have an origin, live, and go extinct. Many of the findings we have described in this chapter clearly show how evolutionary perspectives can provide compelling underlying explicative grounds for contemporary genomic phenomena.

5.6 Technical Note

5.6.1 "Outgroup" Concept

When two species display a difference (in our case a chromosomal difference), it is important to know which of the two forms is ancestral and which is derivative to resolve the polarity of the difference. The solution is to introduce into the analysis of one or multiple closely related species chosen from those that diverged from the common ancestor before the two species under study. More technically, an outgroup species is defined as species or group of species closely related to but not included within the taxon.

5.6.2 Synteny Studies Exploiting BAC or Fosmid Clones in FISH Experiments

The conspicuous number of mapped human clones, as can be graphically seen in genome browsers (see the track "BAC End Pairs" in UCSC, for instance), is a side effect of the hierarchical approach utilized to sequence the human genome. As a first step toward sequencing, a very large number of BAC clones were ordered in contigs by characterizing their STS content, by fingerprinting, and by BAC end sequencing (BES). Then, a minimal number of overlapping BACs (or, occasionally, cosmid clones) were fully sequenced. This subset of clones constituted the "golden path."

Following the completion of the human genome sequencing, all non-sequenced BACs were precisely placed on the sequence itself by BLASTing their BES against the human genome. This was possible only for the subset of BAC clones whose ends were both single copy. The complete set of BES data is present in the "Trace archive" database at the NCBI (http://www.ncbi.nlm.nih.gov/Traces/). Note that the fully sequenced BACs of the "golden path" are not present in the "BAC end pairs" track, but present in the "Clone coverage" and "Assembly from Fragments" tracks

(UCSC) according to their accession number. It is anyway possible to discover the name of the clone that contributed that sequence by querying the accession number at NCBI (http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=search&db=nucleotide). Recently, the ends of several fosmid libraries (~40 kb insert) were sequenced as part of a copy number variation research projects (Kidd et al. 2008; Tuzun et al. 2005). The fosmids of the first library are present in the track "Fosmid End Pairs" of UCSC. Many of these resources are available from the P. de Jong Laboratory (http://bacpac.chori.org/).

Human BAC clones can be successfully FISHed on apes and Old World monkeys. The success rate decreases in New World monkeys. A rule of thumb for sequence homology comparison among species says that it approximately diminishes by 1% every 5 million years of divergence. Hybridization efficiency can be improved by decreasing the hybridization stringency conditions and increasing the hybridization time. Additionally, pools of 2–4 overlapping BACs can be hybridized together, and gene-rich BACs should be preferred, because gene domains can be supposed to be more conserved. At the present, with several mammal genomes sequenced, the evolutionary conservation of a region can be easily checked by visually inspecting the "Conservation" track at UCSC browser.

The genome sequencing of non-human species was usually achieved using a pure shotgun method, which is less time- and money-consuming, but has a higher risk of mis-assembly as compared to the hierarchical approach (Green 1997). The BES pairs of a specific BAC library are usually utilized to improve the shot-gun assembly. As a consequence, a species-specific BAC library is usually available for a sequenced genome. These BACs can be very helpful. Appropriate BAC clones can be identified by their BES, present in the "Trace archive" at the NCBI (see above).

Acknowledgements The MiUR (Ministero della Universita' e della Ricerca) support is acknowledged.

References

Alonso A, Mahmood R, Li S, Cheung F, Yoda K, Warburton PE (2003) Genomic microarray analysis reveals distinct locations for the CENP-A binding domains in three human chromosome 13q32 neocentromeres. Hum Mol Genet 12:2711–2721

Alonso A, Fritz B, Hasson D, Abrusan G, Cheung F, Yoda K, Radlwimmer B, Ladurner AG, Warburton PE (2007) Co-localization of CENP-C and CENP-H to discontinuous domains of CENP-A chromatin at human neocentromeres. Genome Biol (www) 8:R148

Amor DJ, Choo KH (2002) Neocentromeres: role in human disease, evolution, and centromere study. Am J Hum Genet 71:695–714

Amor DJ, Bentley K, Ryan J, Perry J, Wong L, Slater H, Choo KH (2004) Human centromere repositioning "in progress". Proc Natl Acad Sci USA 101:6542–6547

Armengol L, Pujana MA, Cheung J, Scherer SW, Estivill X (2003) Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. Hum Mol Genet 12:2201–2208

Bailey JA, Baertsch R, Kent WJ, Haussler D, Eichler EE (2004) Hotspots of mammalian chromosomal evolution. Genome Biol (www) 5:R23

Bukvic N, Susca F, Gentile M, Tangari E, Iannirubertо A, Guanti G (1996) An unusual dicentric Y chromosome with a functional centromere with no detectable alpha-satellite. Hum Genet 97:453–456

Capozzi O, Purgato S, Verdun di Cantogno L, Grosso E, Ciccone R, Zuffardi O, Della Valle G, Rocchi M (2008) Evolutionary and clinical neocentromeres: two faces of the same coin. Chromosoma 117:339–344

Capozzi O, Purgato S, D'Addabbo P, Archidiacono N, Battaglia P, Baroncini A, Capucci A, Stanyon R, Della Valle G, Rocchi M. Evolutionary descent of a human chromosome 6 neocentromere: a jump back to 17 million years ago. Genome Res (in press)

Carbone L, Ventura M, Tempesta S, Rocchi M, Archidiacono N (2002) Evolutionary history of chromosome 10 in primates. Chromosoma 111:267–272

Carbone L, Nergadze SG, Magnani E, Misceo D, Francesca Cardone M, Roberto R, Bertoni L, Attolini C, Francesca Piras M, de Jong P, Raudsepp T, Chowdhary BP, Guerin G, Archidiacono N, Rocchi M, Giulotto E (2006) Evolutionary movement of centromeres in horse, donkey, and zebra. Genomics 87:777–782

Cardone MF, Alonso A, Pazienza M, Ventura M, Montemurro G, Carbone L, de Jong PJ, Stanyon R, D'Addabbo P, Archidiacono N, She X, Eichler EE, Warburton PE, Rocchi M (2006) Independent centromere formation in a capricious, gene-free domain of chromosome 13q21 in Old World monkeys and pigs. Genome Biol (www) 7:R91

Cardone MF, Lomiento M, Teti MG, Misceo D, Roberto R, Capozzi O, D'Addabbo P, Ventura M, Rocchi M, Archidiacono N (2007) Evolutionary history of chromosome 11 featuring four distinct centromere repositioning events in Catarrhini. Genomics 90:35–43

Chueh AC, Wong LH, Wong N, Choo KHx (2005) Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. Hum Mol Genet 14:85–93

Consortium ICGS (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716

Eder V, Ventura M, Ianigro M, Teti M, Rocchi M, Archidiacono N (2003) Chromosome 6 phylogeny in primates and centromere repositioning. Mol Biol Evol 20:1506–1512

Eichler EE (2001) Segmental duplications: what's missing, misassigned, and misassembled-and should we care. Genome Res 11:653–656

Everts-van der Wind A, Larkin DM, Green CA, Elliott JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA (2005) A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. Proc Natl Acad Sci USA 102:18526–18531

Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, Weber JL, Ledbetter DH, Zuffardi O (2001) Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am J Hum Genet 68:874–883

Green P (1997) Against a whole-genome shotgun. Genome Res 7:410–417

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098–1102

Ishii K, Ogiyama Y, Chikashige Y, Soejima S, Masuda F, Kakuma T, Hiraoka Y, Takahashi K (2008) Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. Science 321:1088–1091

Kasai F, Garcia C, Arruga MV, Ferguson-Smith MA (2003) Chromosome homology between chicken (Gallus gallus domesticus) and the red-legged partridge (Alectoris rufa); evidence of the occurrence of a neocentromere during evolution. Cytogenet Genome Res 102:326–330

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tuzun E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W, Gustafson E, McKernan K, Chen L, Malig M, Smith JD, Korn JM, McCarroll SA,

Altshuler DA, Peiffer DA, Dorschner M, Stamatoyannopoulos J, Schwartz D, Nickerson DA, Mullikin JC, Wilson RK, Bruhn L, Olson MV, Kaul R, Smith DR, Eichler EE (2008) Mapping and sequencing of structural variation from eight human genomes. Nature 453:56–64

Kobayashi T, Yamada F, Hashimoto T, Abe S, Matsuda Y, Kuroiwa A (2008) Centromere repositioning in the X chromosome of XO/XO mammals, Ryukyu spiny rat. Chromosome Res 16:587–593

Larkin DM, Everts-van der Wind A, Rebeiz M, Schweitzer PA, Bachman S, Green C, Wright CL, Campos EJ, Benson LD, Edwards J, Liu L, Osoegawa K, Womack JE, de Jong PJ, Lewin HA (2003) A cattle–human comparative map built with cattle BAC-ends and human genome sequence. Genome Res 13:1996–1972

Lo AW, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E, Magliano DJ, Choo KH (2001a) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. EMBO J 20:2087–2096

Lo AW, Magliano DJ, Sibson MC, Kalitsis P, Craig JM, Choo KH (2001b) A novel chromatin immunoprecipitation and array (cia) analysis identifies a 460-kb cenp-a-binding neocentromere DNA. Genome Res 11:448–457

Lomiento M, Jiang Z, D'Addabbo P, Eichler EE, Rocchi M (2008) Evolutionary-new centromeres preferentially emerge within gene deserts. Genome Biol (www) 9(12):R173

Marshall OJ, Chueh AC, Wong LH, Choo KH (2008) Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am J Hum Genet 82:261–282

Misceo D, Cardone MF, Carbone L, D'Addabbo P, de Jong PJ, Rocchi M, Archidiacono N (2005) Evolutionary history of chromosome 20. Mol Biol Evol 22:360–366

Misceo D, Capozzi O, Roberto R, Dell'Oglio MP, Rocchi M, Stanyon R, Archidiacono N (2008) Tracking the complex flow of chromosome rearrangements from the Hominoidea Ancestor to extant Hylobates and Nomascus Gibbons by high-resolution synteny mapping. Genome Res 18:1530–1537

Montefalcone G, Tempesta S, Rocchi M, Archidiacono N (1999) Centromere repositioning. Genome Res 9:1184–1188

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'Brien SJ, Pevzner PA, Lewin HA (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309:613–617

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145

O'Neill RJ, Eldridge MD, Metcalfe CJ (2004) Centromere dynamics and chromosome evolution in marsupials. J Hered 95:375–381

Oakenfull EA, Clegg JB (1998) Phylogenetic relationships within the genus Equus and the evolution of alpha and theta globin genes. J Mol Evol 47:772–783

Oakenfull E, Lim H, Ryder O (2000) A survey of equid mitochondrial DNA: Implications for the evolution, genetic diversity and conservation of *Equus*. Conservation Genet 1:341–355

Opazo JC, Wildman DE, Prychitko T, Johnson RM, Goodman M (2006) Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). Mol Phylogenet Evol 40:274–280

Pardo-Manuel de Villena F, Sapienza C (2001) Transmission ratio distortion in offspring of heterozygous female carriers of Robertsonian translocations. Hum Genet 108:31–36

Pevzner P, Tesler G (2003) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc Natl Acad Sci USA 100:7672–7677

Raaum RL, Sterner KN, Noviello CM, Stewart CB, Disotell TR (2005) Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. J Hum Evol 48:237–257

Rivera H, Vassquez AI, Ayala-Madrigal ML, Ramirez-Duenas ML, Davalos IP (1996) Alphoidless centromere of a familial unstable inverted Y chromosome. Ann Genet 39:236–239

Roberto R, Capozzi O, Wilson RK, Mardis ER, Lomiento M, Tuzun E, Cheng Z, Mootnick AR, Archidiacono N, Rocchi M, Eichler EE (2007) Molecular refinement of gibbon genome rearrangement. Genome Res 17:249–257

Roberto R, Misceo D, D'Addabbo P, Archidiacono N, Rocchi M (2008) Refinement of macaque synteny arrangement with respect to the official rheMac2 macaque sequence assembly. Chromosome Res 16(7):977–985

Rocchi M, Archidiacono N, Stanyon R (2006) Ancestral genomes reconstruction: An integrated, multi-disciplinary approach is needed. Genome Res 16:1441–1444

Saffery R, Sumer H, Hassan S, Wong LH, Craig JM, Todokoro K, Anderson M, Stafford A, Choo KH (2003) Transcription within a functional human centromere. Mol Cell 12:509–516

She X, Horvath JE, Jiang Z, Liu G, Furey TS, Christ L, Clark R, Graves T, Gulden CL, Alkan C, Bailey JA, Sahinalp C, Rocchi M, Haussler D, Wilson RK, Miller W, Schwartz S, Eichler EE (2004) The structure and evolution of centromeric transition regions within the human genome. Nature 430:857–864

Stanyon R, Rocchi M, Capozzi O, Roberto R, Misceo D, Ventura M, Cardone M, Bigoni F, Archidiacono N (2008) Primate chromosome evolution: ancestral karyotypes, marker order and neocentromeres. Chromosome Res 16:17–39

Sumer H, Craig JM, Sibson M, Choo KH (2003) A rapid method of genomic array analysis of scaffold/matrix attachment regions (S/MARs) identifies a 2.5-Mb region of enhanced scaffold/matrix attachment at a human neocentromere. Genome Res 13:1737–1743

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE (2005) Fine-scale structural variation of the human genome. Nat Genet 37:727–732

Tyler-Smith C, Gimelli G, Giglio S, Floridia G, Pandya A, Terzoli G, Warburton PE, Earnshaw WC, Zuffardi O (1999) Transmission of a fully functional human neocentromere through three generations. Am J Hum Genet 64:1440–1444

Ventura M, Archidiacono N, Rocchi M (2001) Centromere emergence in evolution. Genome Res 11:595–599

Ventura M, Mudge JM, Palumbo V, Burn S, Blennow E, Pierluigi M, Giorda R, Zuffardi O, Archidiacono N, Jackson MS, Rocchi M (2003) Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. Genome Res 13:2059–2068

Ventura M, Weigl S, Carbone L, Cardone MF, Misceo D, Teti M, D'Addabbo P, Wandall A, Björck E, de Jong P, She X, Eichler EE, Archidiacono N, Rocchi M (2004) Recurrent sites for new centromere seeding. Genome Res 14:1696–1703

Ventura M, Antonacci F, Cardone MF, Stanyon R, D'Addabbo P, Cellamare A, Sprague LJ, Eichler EE, Archidiacono N, Rocchi M (2007) Evolutionary formation of new centromeres in macaque. Science 316:243–246

Villasante A, Abad JP, Mendez-Lago M (2007) Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. Proc Natl Acad Sci USA 104:10542–10547

Warburton PE (2004) Chromosomal dynamics of human neocentromere formation. Chromosome Res 12:617–626

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe

DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

Wong LH, Choo KH (2001) Centromere on the move. Genome Res 11:513–516

Table 5.1 ENC in a *red circle* indicates the emergence of an evolutionary new centromere. The *red arrows* point to inactivated centromeres. For details see text
Supplement Table 5.1 BAC clones used to delineate evolutionary history of chromosomes in primates

		HSA1			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-421C4	BES	chr1:1,247,484-1,432,829	1p36.33		
RP11-265F14	AL512883	chr1:15,630,693-15,735,006	1p32.21		
RP11-266K22	AL451070	chr1:31,512,242-31,645,345	1p35.2		
RP5-1154B21	(D1S3315)	chr1:47,605,724-47,806,009	1p33		
RP11-55M23	BES	chr1:55,214,018-55,384,910	1p32.3		
				LLA9ENC	
RP11-316C12	AL627317	chr1:71,621,580-71,715,216	1p31.1		
RP11-254E16	BES	chr1:84,539,796-84,689,450	1p31.1		
RP11-138K16	AC093559	chr1:99,728,340-99,904,318	1p21.2		
RP11-284N8	AL365361	chr1:110,897,276-111,090,458	1p13.3		
				LLA28ENC	Stanyon et al. (2008)
RP11-192J8	BES	chr1:118,333,066-118,333,600	1p12		
RP5-1042I8	AL359752	chr1:120,134,618-120,272,572	1p12		
CEN				APCEN	Stanyon et al. (2008)
HETEROCHRO	MATIN				
RP11-35B4	AL359093	chr1:143,999,789-144,166,609	1q21.1+…		
RP11-98F1	AL353760	chr1:153,539,660-153,543,654	1q22		
RP11-8D14	AC068728	chr1:158,524,306-159,027,864	1q23.3		
RP11-117F19	BES	chr1:160,616,094-160,786,583	1q23.3		
RP11-331H2	AL392003	chr1:161,033,791-161,225,664	1q23.3		
RP11-593N18	BES	chr1:165,733,633-165,896,797	1q24.2		
RP11-332H17	AL356475	chr1:168,242,366-168,350,412	1q24.2		
RP11-170H10	BES	chr1:177,378,863-177,558,311	1q25.2		
RP11-152A16	BES	chr1:177,339,824-177,521,905	1q25.2		
RP11-46A10	BES	chr1:179,093,659-179,296,257	1q25.3		
RP11-453M18	BES	chr1:179,309,603-179,489,310	1q25.3		

(continued)

Table 5.1 (continued)

		HSA1			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-382D12	AL445228	chr1:182,516,029-182,632,280	1q25.3		
RP11-134C1	BES	chr1:186,252,053-186,417,889	1q31.1		
RP11-92O4	BES	chr1:186,371,022-186,536,053	1q31.1		
RP11-13G5	BES	chr1:190,775,806-190,935,812	1q31.2		
RP11-173E24	AL138926	chr1:193,522,550-193,691,711	1q31.3		
RP11-44M20	BES	chr1:194,486,097-194,633,832	1q31.3		
RP11-112O19	BES	chr1:194,850,927-195,161,754	1q31.3		
RP11-192O22	BES	chr1:195,224,762-195,369,931	1q31.3		
RP11-553K8	AL157402	chr1:196,750,597-196,960,927	1q31.3		
RP11-57I17	AL137789	chr1:205,811,787-205,957,118	1q32.2		
RP11-2P2	BES	chr1:206,365,585-206,526,445	1q32.2		
RP11-167J2	BES	chr1:207,651,871-207,813,996	1q32.2		
RP11-345I23	BES	chr1:207,907,102-208,088,214	1q32.2		
RP11-237I23	BES	chr1:208,074,198-208,253,751	1q32.2		
RP11-168F20	BES	chr1:208,231,193-208,397,063	1q32.2		
RP11-123O6	BES	chr1:209,402,692-209,548,350	1q32.2-1q32.3		
RP11-74E6	BES	chr1:212,560,677-212,729,794	1q32.3-1q41		
RP11-324K19	AC118468	chr1:219,203,440-219,218,989	1q41		
RP11-351B5	BES	chr1:220,760,290-220,938,014	1q41		
RP11-18A13	BES	chr1:221,338,590-221,505,250	1q41		
				CJA19ENC	Unpublished data
RP11-122D22	BES	chr1:223,625,059-223,799,771	1q42.12		
RP11-3K22	AL359874	chr1:227,082,837-227,162,782	1q42.13		
RP11-108F13	BES	chr1:227,804,416-227,983,307	1q42.13		
RP11-543E8	FISH	chr1:228,724,827-228,906,001	1q42.13-42.2		
RP11-933K5	BES	chr1:228,896,382-229,080,653	1q42.2		
RP11-316N16	BES	chr1:229,077,217-229,258,402	1q42.2		

RP11-281B4	BES	chr1:229,444,660-229,608,742	1q42.2		
RP11-88N18	BES	chr1:229,794,597-229,951,130	1q42.2		
RP11-210E16	BES	chr1:230,558,779-230,738,335	1q42.2		
RP11-155C15	BES	chr1:232,687,473-232,687,928	1q42.2		
RP11-385F5	AL359921	chr1:234,752,824-234,966,581	1q43		
RP11-438F14	AC098483	chr1:246,754,133-246,932,000	1q44		

		HSA3			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-151A4(A)	BES	chr3:636,173-795,419	3p26.3		
RP11-183N22	AL512885	chr3:4,328,222-4,493,696	3p26.1		
RP11-48N24	BES	chr3:7,397,489-7,541,994	3p26.1		
RP11-732C9	BES	chr3:12,441,757-12,649,037	3p25.2		
RP11-316A10	AC090937	chr3:14,886,290-15,046,968	3p25.1		
RP11-616M11(B)	AC090954	chr3:15,045,785-15,213,797	3p25.1		
RP11-421B21(C)	AC090949	chr3:15,147,209-15,324,532	3p25.1		
RP11-109D5	BES	chr3:25,497,576-25,697,390	3p24.2		
			3p23	CNC	Maraschio et al. (1996)
RP11-627J17	AC112211	chr3:32,980,609-33,163,117	3p23		
RP11-240N7	BES	chr3:36,506,239-36,658,135	3p22.2		
RP11-607P24	BES	chr3:36,298,506-36,506,070	3p22.3-22.2		
RP11-491D6	AC006583	chr3:37,034,150-37,136,520	3p22.3		
RP11-713K14	AP006242	chr3:37,868,651-38,036,022	3p22.3		
RP11-409G11	BES	chr3:39,926,773-40,069,893	3p22.1		
RP11-465K13	BES	chr3:40,934,090-41,113,356	3p22.1		
RP11-756A10	AC099059	chr3:41,445,167-41,603,742	3p22.1		
RP11-626A1	AC137935	chr3:41,872,546-42,048,517	3p22.1		
RP11-12I10	BES	chr3:42,489,150-42,664,465	3p22.1		
RP11-1047D9	BES	chr3:42,688,112-42,792,912	3p22,1		
RP11-625B23	BES	chr3:43,135,351-43,328,084	3p22.1		

(continued)

Table 5.1 (continued)

		HSA3			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-353H3(D)	BES	chr3:43,377,484-43,547,690	3p22.1		
				CAE22ENC	Ventura et al. (2004)
RP11-395P16(E)	AC130472	chr3:47,584,747-47,778,906	3p21.31		
RP11-380J21	AC136275	chr3:64,182,355-64,200,696	3p14.1		
RP11-151M23	BES	chr3:67,043,727-67,181,276	3p14.1		
RP11-158P4	BES	chr3:73,574,165-73,731,870	3p13		
RP11-634L22(F)	BES	chr3:75,452,260-75,628,601	3p12.3		
				CJA21ENC	Ventura et al. (2004)
RP11-180C9(G)	BES	chr3:75,997,245-76,170,439	3p12.3		
RP11-536K4	AC016942	chr3:76,682,896-76,834,909	3p12.3		
RP11-655A17	BES	chr3:87,099,698-87,270,490	3p11.2-12.1		
RP11-547K2(H)	AC107028	chr3:89,543,932-89,670,647	3p11.1		
CEN		chr3:89,700,001-93,200,000		ANTHROPOIDEA ENC	Ventura et al. (2004)
RP11-124L3(I)	BES	chr3:94,987,545-95,112,295	3q11.2		
RP11-91M15	BES	chr3:96,443,224-96,627,928	3q11.2		
RP11-117C10	BES	chr3:99,602,287-99,735,761	3q11.2		
RP11-454H13	AC084198	chr3:102,798,688-102,993,642	3q12.3		
RP11-305I9	AC092981	chr3:120,465,181-120,625,176	3q13.32-13.33		
RP11-757I12	AC092908	chr3:123,727,494-123,901,752	3q21.1		
RP11-257B7	BES	chr3:125,091,060-125,253,000	3q21.1		
RP11-98E19	BES	chr3:126,496,462-126,666,036	3q21.2		
RP11-26M12(J)	BES	chr3:130,089,343-130,276,597	3q21.3		
RP11-787P10(K)	BES	chr3:131,347,364-131,500,312	3p21.3-22.1		
RP11-21N8	BES	chr3:131,810,630-131,961,198	3q22.1		
RP11-58H13	BES	chr3:134,587,560-134,765,508	3q22.1		
RP11-45B17	BES	chr3:139,765,789-139,942,541	3q22.3		
RP11-13N24	BES	chr3:145,950,523-146,100,894	3q24		

RP11-505J9	BES	chr3:149,845,223-150,049,901	3q24		
			3q24	HRC	Ventura et al. (2004)
RP11-36G5	BES	chr3:151,686,443-151,861,594	3q25.1		
RP11-484J9	BES	chr3:154,596,424-154,771,945	3q25.2		
RP11-142B1	BES	chr3:162,044,658-162,223,487	3q26.1		
			3q26.1	CNC	Ventura et al. (2004)
RP11-498P15	AC112906	chr3:163,530,323-163,648,093	3q26.1		
RP11-355I21	AC025826	chr3:163,822,353-164,122,697	3q26.1		
RP11-498P15	AC112906	chr3:163,530,323-163,648,093	3q26.1		
RP11-355I21(L)	AC025826	chr3:163,822,353-164,122,697	3q26.1		
				OWMENC	Ventura et al. (2004)
RP11-418B12(M)	AC079910	chr3:164,539,721-164,707,127	3q26.1		
RP11-526M23	AC048332	chr3:166,898,263-167,089,798	3q26.1		
RP11-114M1	BES	chr3:178,755,562-178,913,002	3q26.32		
RP11-121O16(N)	BES	chr3:179,246,025-179,381,716	3q26.32		
				CJA15,LLA22, CJA17ENCs	Ventura et al. (2004)
RP11-160P8(O)	BES	chr3:179,811,584-179,969,664	3q26.32		
RP11-102M21	BES	chr3:180,795,920-180,965,055	3q26.33		
RP11-218A22	AC108670	chr3:186,915,322-187,078,478	3q27.2		
			3q27	CNC	Papenhausen et al. (1995)
RP11-709J22	BES	chr3:187,553,921-187,754,882	3q27.3		
RP11-42D20	AC007690	chr3:188,091,027-188,272,557	3q27.3		
RP11-298A18	AC063932	chr3:189,731,862-189,918,758	3q28		
RP11-153K2	BES	chr3:192,086,117-192,227,844	3q28		
RP11-6E10	BES	chr3:197,556,002-197,719,622	3q29		
RP11-313F11(P)	FISH	chr3:198,844,624-198,845,224	3q29		
				CMO20ENC	Ventura et al. (2004)

(continued)

Table 5.1 (continued)

		HSA4			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-61B7	BES	chr4:39,428-230,148	4p16.3		
RP11-167K22	BES	chr4:15,166,060-15,327,998	4p15.32-15.33		
RP11-102K4	BES	chr4:18,302,511-18,440,696	4p15.32		
RP11-585D5	BES	chr4:21,855,149-22,058,506	4p15.31		
RP11-156A17	BES	chr4:21,620,803-21,774,876	4p15.31		
RP11-362I16	AC093814	chr4:22,002,309-22,165,509	4p15.31		
RP11-157B23	BES	chr4:23,884,053-24,050,372	4p15.2		
RP11-125D22	BES	chr4:26,121,547-26,273,493	4p15.2		
RP11-100L2	BES	chr4:28,768,899-28,923,065	4p15.1		
RP11-164K20	BES	chr4:29,841,543-30,016,485	4p15.1		
RP11-124E24	BES	chr4:32,539,454-32,710,916	4p15.1		
RP11-135M12	AC096735	chr4:35,206,808-35,400,317	4p15		
RP11-108H14	BES	chr4:36,852,417-37,053,575	4p14		
RP11-103K10	BES	chr4:38,660,498-38,852,024	4p14		
RP11-473D12	AC108149	chr4:43,753,530-43,839,853	4p13		
RP11-317G22	AC020593	chr4:48,589,290-48,773,495	4p12		
CEN					
RP11-365H22	AC027271	chr4:52,354,875-52,532,859	4q11		
RP11-669F1	BES	chr4:68,763,265-68,894,804	4q13.2		
			4q21.1-21.3	CNC	Grimbacher et al. (1999); Warburton et al. (2003)
RP11-458G13	BES	chr4:87,075,729-87,246,734	4q21.23-21.3		
RP11-209G6	BES	chr4:88,092,435-88,250,275			
				HRC	Amor et al. (2004)

RP11-204I22	BES	chr4:89,693,568-89,857,513			
RP11-499E18	AC098487	chr4:103,434,684-103,598,599	4q24		
				LLA19, LLA4ENC	Stanyon et al. (2008)
RP11-510D4	AC092661	chr4:118,641,765-118,817,790	4q26		
RP11-45L5	AC015631	chr4:135,278,394-135,423,531	4q28.3		
RP11-780M14	AC104090	chr4:144,875,612-144,976,649	4q31.21		
RP11-663M18	AC109823	chr4:159,961,327-160,041,534	4q32.1		
RP11-493C20	AC098867	chr4:164,688,088-164,830,401	4q32.3		
RP11-808H17	AC079240	chr4:165,338,176-165,541,022	4q32.3		
RP11-443J23	AC093842	chr4:166,667,434-166,780,458	4q32.3		
RP11-511B7	AC080079	chr4:166,778,459-166,890,974	4q32.3		
RP11-371E22	AC097507	chr4:167,055,229-167,222,260	4q32.3		
RP11-624O16	AC093874	chr4:167,220,261-167,378,505	4q32.3		
RP11-436G13	AC107055	chr4:167,376,506-167,521,299	4q32.3		
RP11-368M2	BES	chr4:167,334,183-167,519,234	4q32.3		
RP11-13P1	BES	chr4:167,405,757-167,570,823	4q32.3		
RP11-662N23	BES	chr4:167,405,729-167,578,663	4q32.3		
				NWMCEN	Stanyon et al. (2008)
RP11-455K3	BES	chr4:167,510,402-167,693,775	4q32.3		
RP11-638N11	BES	chr4:167,724,651-167,927,839	4q32.3		
RP11-662D13	AC068989	chr4:169,331,047-169,528,540	4q32.3		
RP11-648O9	AC106878	chr4:170,815,533-170,950,874	4q33		
RP11-51M24	BES	chr4:175,426,843-175,581,124	4q34.1		
RP11-99E17	BES	chr4:180,594,112-180,767,207	4q34.3		
RP11-104E20	BES	chr4:185,808,872-185,994,898	4q35.1		
RP11-45C13	BES	chr4:187,655,959-187,810,778	4q35.2		
RP11-138B4	BES	chr4:188,409,307-188,555,694	4q35.2		
RP11-242B20	BES	chr4:190,768,122-190,931,965	4q35.2		

(continued)

Table 5.1 (continued)

		HSA5			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-58A5	BES	chr5:4,965,694-5,123,154	5p15.32		
RP11-5N8	BES	chr5:14,926,850-15,108,182	5p15.2		
			5p14-p15.1	CNC	Fritz et al. (2001)
RP11-12C2	AC114298	chr5:23,056,186-23,144,972	5p14.3		
RP11-94E6	BES	chr5:33,737,270-33,926,014	5p13.3		
RP11-159F24	BES	chr5:43,509,883-43,672,607	5p12		
CEN					
RP11-160F8	BES	chr5:53,365,254-53,520,258	5q11.2		
RP11-298P6	AC109465	chr5:64,070,008-64,258,158	5q12.3		
RP11-172K14	BES	chr5:74,490,650-74,673,873	5q13.3		
				CMO11-CMO14CEN	Stanyon et al. (2008)
RP11-258M21	BES	chr5:85,443,998-85,587,282	5q14.3		
RP11-297G19	AC093268	chr5:93,288,262-93,463,665	5q15		
				LLA3ENC	Stanyon et al. (2008)
RP11-326M11	BES	chr5:105,194,174-105,355,876	5q21.3		
RP11-81C5	BES	chr5:115,183,047-115,366,698	5q23.1		
RP11-209F21	BES	chr5:124,786,541-124,969,874	5q23.2		
RP11-42M12	BES	chr5:127,176,898-127,327,842	5q23.2		
RP11-186F1	BES	chr5:130,344,505-130,522,404	5q31.1		
RP11-4E3	BES	chr5:133,085,104-133,272,697	5q31.1		
RP11-1030O9	BES	chr5:133,318,962-133,507,753	5q31.1		
RP11-737P20	BES	chr5:133,455,828-133,650,405	5q31.1		
RP11-21C10	BES	chr5:133,880,133-134,045,963	5q31.1		
RP11-114H21	BES	chr5:135,739,999-135,916,051	5q31.2		
RP11-365D10	BES	chr5:144,529,859-144,719,618	5q32		
RP11-170L13	BES	chr5:155,123,977-155,288,472	5q33.2		
RP11-367N22	BES	chr5:156,258,929-156,421,855	5q33.3		

RP11-52L13	BES	chr5:156,420,347-156,582,237	5q33.3		
RP11-92E20	BES	chr5:156,594,140-156,769,665	5q33.3		
RP11-631N12	BES	chr5:157,000,006-157,172,852	5q33.3		
RP11-82E8	BES	chr5:157,616,481-157,787,918	5q33.3		
RP11-678F4	BES	chr5:158,490,847-158,671,405	5q33.3		
RP11-90N23	FISH	chr5:159,983,909-159,984,761	5q34		
RP11-114D4	BES	chr5:160,424,255-160,577,512	5q34		
RP11-569B13	AC091984	chr5:161,495,046-161,702,085	5q34		
RP11-88J19	BES	chr5:162,047,143-162,237,277	5q34		
RP11-653G7	BES	chr5:163,166,056-163,341,883	5q34		
RP11-308N24	AC109466	chr5:164,314,289-164,468,103	5q34		
RP11-90C21	BES	chr5:165,262,926-165,426,112	5q34		
RP11-436K21	BES	chr5:166,079,437-166,247,339	5q34		
RP11-69K7	BES	chr5:167,097,350-167,258,685	5q34		
RP11-14K9	BES	chr5:168,358,455-168,532,236	5q35.1		
				LLA11CEN	Stanyon et al. (2008)
RP11-170N13	BES	chr5:168,433,613-168,593,223	5q35.1		
RP11-270N4	BES	chr5:168,532,261-168,704,355	5q35.1		
RP11-486H5	BES	chr5:168,909,435-169,092,364	5q35.1		
RP11-15F10	BES	chr5:169,073,440-169,267,747	5q35.1		
RP11-117L6	BES	chr5:170,679,528-170,854,638	5q35.1		
RP11-48K2	BES	chr5:172,952,607-173,131,912	5q35.2		
RP11-125L2	BES	chr5:173,447,703-173,616,629	5q35.2		
RP11-298C7	BES	chr5:176,032,557-176,197,535	5q35.2		
RP11-452O4	BES	chr5:177,234,210-177,410,189	5q35.3		
				CJA2-SSC20-SSC1CEN	Stanyon et al. (2008)

		HSA6			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-328C17	AL365272	chr6:213,636-346,084	6p25.3		

(continued)

Table 5.1 (continued)

		HSA6			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-391F23	AL589203	chr6:929,025-940,528	6p25.3		
RP11-125I8	BES	chr6:10,001,499-10,140,467	6p24.3		
RP11-15I14	BES	chr6:10,459,005-10,639,794	6p24.3		
RP11-147C6	BES	chr6:10,622,819-10,774,675	6p24.2		
RP11-48D18	BES	chr6:10,937,982-11,123,046	6p24.2		
RP11-4A24	AL137221	chr6:12,238,011-12,244,433	6p24.1		
RP11-27M22	BES	chr6:14,673,843-14,829,605	6p23		
RP11-61I16	BES	chr6:15,365,240-15,519,994	6p23		
RP11-17L3	BES	chr6:15,716,005-15,896,106	6p22.3		
RP11-90O12	BES	chr6:16,445,725-16,630,424	6p22.3		
RP11-59N15	BES	chr6:26,015,628-26,168,053	6p22.2		
		chr6:26,407,000-26,491,000		HRC(ChIP-on-chip)	Capozzi et al. (2008a)
RP11-911D8	BES	chr6:27,069,535-27,253,168	6p22.1		
RP11-297M4	BES	chr6:29,016,624-29,189,711	6p22.1		
RP11-99D3	BES	chr6:29,049,490-29,222,060	6p22.1		
				ANCESTRALCEN.	Capozzi et al. (2008a)
RP11-261L19	BES	chr6:29,259,359-29,405,414	6p22.1		
RP11-751N3	BES	chr6:29,555,726-29,748,946	6p22.1		
RP11-351O4	BES	chr6:30,258,900-30,456,684	6p21.33		
RP11-1021F13	BES	chr6:30,304,165-30,524,743	6p21.33		
RP11-349M22	BES	chr6:30,802,793-30,972,043	6p21.33		
RP11-754H10	BES	chr6:33,960,388-34,137,229	6p21.31		
RP11-61E9	BES	chr6:34,202,278-34,379,674	6p21.31		
RP11-481A14	BES	chr6:34,820,453-34,979,774	6p21.31		
RP11-615A19	BES	chr6:34,993,942-35,155,751	6p21.31		
RP11-10I8	BES	chr6:39,087,900-39,254,420	6p21.2		
RP11-25I3	BES	chr6:41,836,819-42,004,320	6p21.1		

RP1-139D8	AL096814	chr6:42,208,853-42,375,930	6p21.1		
RP11-397G17	BES	chr6:50,026,694-50,190,757	6p12.3		
RP11-346L9	BES	chr6:57,351,232-57,548,984	6p11.2		
RP11-791F20	BES	chr6:57,500,124-57,690,152	6p11.2		
RP11-343D24	BES	chr6:57,644,937-57,835,610	6p11.2		
RP11-799H20	BES	chr6:57,787,081-58,000,708	6p11.2		
RP11-484F20	BES	chr6:58,720,610-58,883,743	6p11.1		
CEN				HOMINOIDEA ENC	Eder et al. (2003)
RP11-346M3	BES	chr6:62,456,388-62,630,578	6q11.1		
RP11-474L11	BES	chr6:76,244,412-76,429,104	6q14.1		
RP3-494K13	AL136312	chr6:85,740,159-85,796,186	6q14.3		
RP11-451P21	BES	chr6:96,988,167-97,146,901	6q16.1		
RP11-117A20	AL589920	chr6:119,888,999-119,906,826	6q22.31		
RP11-472E5	AL138828	chr6:136,464,198-136,605,737	6q23.3		
RP11-478J9	BES	chr6:140,333,714-140,416,207	6q24.1		
				OWMENC	Ventura et al. (2007)
RP11-474A9	BES	chr6:145,651,644-145,845,896	6q24.3		
RP11-64M7	AL589705	chr6:149,289,814-149,303,728	6q25.1		
			6q26	CNC	Sala et al. (2005)
RP1-230L10	AL137005	chr6:164,038,658-164,142,336	6q27		
RP11-37D8	BES	chr6:168,661,593-168,825,471	6q27		
RP11-302L19	AL596442	chr6:170,264,380-170,375,196	6q27		

		HSA7			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
				CJA2ENC; LLA11ENC	Unpublished data
RP11-713A20	AC093686	chr7:106,471-298,664	7p22.3		
RP11-416J17	AC069288	chr7:1,911,784-2,057,495	7p22.3		
RP11-792G24	BES	chr7:2,339,107-2,562,885	7p22.2		
RP11-400E7	BES	chr7:2,600,022-2,778,338	7p22.2		

(continued)

Table 5.1 (continued)

		HSA7			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-96L18	BES	chr7:2,825,887-2,981,935	7p22.2		
RP11-166P10	BES	chr7:3,369,663-3,531,934	7p22.2		
RP11-160E17	BES	chr7:4,751,001-4,913,015	7p22.1		
RP11-1080O3	BES	chr7:6,392,079-6,613,748	7p22.1		
RP11-1119G2	BES	chr7:6,991,152-7,134,725	7p22.1		
RP11-1061P7	BES	chr7:7,043,428-7,227,820	7p22.1		
RP4-755G17	AC004879	chr7:10,151,763-10,286,666	7p21.3		
RP11-486P11	AC007001	chr7:20,042,179-20,150,596	7p15.3		
RP11-112E16	BES	chr7:30,108,214-30,275,845	7p15.1		
RP11-585N13	BES	chr7:31,423,410-31,589,333	7p15.1		
RP11-714H18	BES	chr7:31,716,263-31,853,656	7p15.1		
RP11-638B17	BES	chr7:32,687,412-32,896,865	7p14.3		
RP11-420P20	BES	chr7:40,248,240-40,427,560	7p14.1		
RP11-653O17	AC073424	chr7:48,207,950-48,399,090	7p12.3		
RP11-339F13	AC073324	chr7:55,222,879-55,348,131	7p11.2		
CEN					
RP11-72B17	BES	chr7:65,153,654-65,319,685	7q11.21		
RP11-105P18	BES	chr7:68,481,085-68,640,842	7q11.22		
RP5-1102A12	AC004963	chr7:70,212,578-70,386,204	7q11.22		
RP11-243I17	BES	chr7:75,622,228-75,783,726	7q11.23		
RP11-982E3	BES	chr7:76,687,499-76,879,221	7q11.23		
RP11-580C19	BES	chr7:83,150,102-83,328,226	7q21.11		
RP11-215P16	AC006036	chr7:90,317,020-90,473,859	7q21.13		
RP11-908F6	BES	chr7:97,256,389-97,437,527	7q21.3		
RP11-150J17	BES	chr7:97,536,166-97,711,886	7q21.3		
RP11-163E9	BES	chr7:101,687,461-101,859,446	7q22.1		
RP11-803J14	BES	chr7:101,984,409-102,291,257	7q22.1		

RP11-282M13	BES	chr7:102,291,028-102,457,862	7q22.1		
RP11-418B19	AC073208	chr7:103,221,699-103,293,304	7q22.1		
RP11-328M22	AC018464.9	chr7:112,279,363-112,435,224	7q31.1		
RP11-22K23	BES	chr7:115,242,134-115,391,679	7q31.2		
RP11-108L6	BES	chr7:116,588,336-116,756,666	7q31.2		
RP11-55P11	BES	chr7:119,167,923-119,349,966	7q31.31		
RP11-3L10	BES	chr7:120,824,541-120,989,241	7q31.32		
RP11-329I5	AC018642.7	chr7:130,598,383-130,792,905	7q32.3		
RP5-839B19	AC006347	chr7:140,157,573-140,227,347	7q34		
RP11-422E4	AC024730.7	chr7:153,750,370-153,901,567	7q36.2		
RP11-764O12	AC006476	chr7_random:1-112,804			

		HSA8			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-18D5	AC090135	chr8:381,182-484,890	8p23.3		
RP11-59B16	BES	chr8:5,798,863-5,947,994	8p23.2		
RP11-737E8	BES	chr8:11,580,455-11,789,912	8p23.1		
RP11-247B12	BES	chr8:11,819,908-11,980,152	8p23.1		
RP11-98O19	BES	chr8:12,259,223-12,433,476	8p23.1		
RP11-45O16	BES	chr8:12,919,224-13,073,779	8p22		
RP11-460L9	BES	chr8:19,538,527-19,705,748	8q21.3		
RP11-583M2	AC051642	chr8:23,420,721-23,595,169	8q21.2		
RP11-120K21	BES	chr8:25,830,685-25,965,011	8q21.2		
RP11-51H24	BES	chr8:30,654,547-30,835,886	8p12		
RP11-10D7	AC013603	chr8:33,487,657-33,665,495	8p12		
RP11-262I23	BES	chr8:39,846,706-40,045,213	8p11.22		
CEN					
RP11-1134I14	BES	chr8:48,063,873-48,241,291	8q11.1-q11.21		
RP11-80E22	BES	chr8:52,787,115-52,932,670	8q11.22-q11.23		
RP11-151B2	BES	chr8:56,450,221-56,608,976	8q12.1		

(continued)

Table 5.1 (continued)

		HSA8			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-36P16	BES	chr8:60,277,203-60,460,134	8q12.1		
RP11-45G14	BES	chr8:62,325,810-62,478,435	8q12.2-q12.3		
RP11-280G9	BES	chr8:62,716,900-62,859,102	8q12.3		
				LLA7ENC	Stanyon et al. (2008)
RP11-382J12	AC022731	chr8:71,614,507-71,778,503	8q13.3		
RP11-75P23	BES	chr8:72,585,391-72,768,280	8q13.3		
RP11-232D14	BES	chr8:73,793,280-73,983,779	8q13.3		
RP11-361C12	BES	chr8:74,618,561-74,774,512	8q21.11		
RP11-300E4	AC100782	chr8:76,034,751-76,219,873	8q21.11		
RP11-706J10	BES	chr8:77,470,475-77,644,774	8q21.11		
RP11-91P17	AC084706	chr8:79,158,450-79,305,261	8q21.12		
RP11-14D5	BES	chr8:86,064,185-86,255,578	8q21.2		
RP11-353O11	AC091184	chr8:90,077,451-90,220,326	8q21.3		
RP11-703K20	BES	chr8:90,220,321-90,398,370	8q21.3		
RP11-179G18	BES	chr8:90,294,331-90,433,195	8q21.3		
RP11-18K20	AC099816	chr8:90,469,419-90,620,876	8q21.3		
RP11-15J4	BES	chr8:92,022,248-92,211,470	8q21.3		
RP11-14G13	BES	chr8:96,166,084-96,341,103	8q22.1		
RP11-122P10	BES	chr8:97,321,450-97,488,017	8q22.1		
RP11-452M24	BES	chr8:98,106,600-98,303,112	8q22.1		
RP11-35A21	BES	chr8:98,181,249-98,335,203	8q22.1		
RP11-828L5	BES	chr8:98,534,473-98,760,672	8q22.1		
RP11-958K24	BES	chr8:98,760,678-98,948,597	8q22.1		
RP11-640O15	BES	chr8:99,070,433-99,228,183	8q22.1-q22.2		
RP11-410L14	AC104986	chr8:99,944,884-100,098,300	8q22.2		
RP11-697C18	AC024996	chr8:113,395,877-113,573,740	8q23.3		
RP11-269I24	AC090987	chr8:131,641,435-131,795,238	8q24.21		

RP11-349C2	AC087337	chr8:145,586,068-145,770,875	8q24.3		
RP4-698E23	AF186192	chr8:145,807,985-145,953,950	8q24.3		
				CMO17ENC	Stanyon et al. (2008)

		HSA9			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-59O6	BES	chr9:188,713-373,816			
RP11-130C19	AL136979	chr9:615,148-812,246	9p24.3		
RP11-341G2	BES	chr9:1,121,123-1,241,689			
RP11-472F14	BES	chr9:6,427,961-6,601,707	9p24.1		
RP11-77E14	AL354694	chr9:7,671,919-7,825,210	9p24.1		
RP11-44k8	BES	chr9:10,913,827-11,089,825	9p23		
	BES	~chr9:10,913,827-11,341,974	9p23	CNC	Satinover et al. (2001)
RP11-23D5	BES	chr9:11,170,427-11,341,974	9p23		
RP11-115I23	BES	chr9:13,017,706-13,186,522	9p23	tumor	Italiano et al. (2006)
RP11-58K1	BES	chr9:15,874,140-16,051,195	9p22.3		
RP11-340N12	FISH	chr9:17,136,369-17,298,494	9p22.2		
RP11-57I14	BES	chr9:19,650,027-19,797,139	9p22.1		
RP11-393P6	AL513317	chr9:23,950,338-24,092,705	9p21.3		
RP11-1006E22	BES	chr9:27,142,243-27,331,367	9p21.2		
RP11-976P13	BES	chr9:30,838,876-31,023,309	9p21.1		
RP11-562M8	AL353717	chr9:32,871,544-32,992,078	9p21.1		
RP11-58A20	BES	chr9:36,392,186-36,539,166	9p13.2		
RP11-3J10	AL138752	chr9:37,745,972-37,935,175	9p13.2		
RP11-168J7	BES	chr9:38,261,095-38,421,467	9p13.1		
RP11-788E5	BES	chr9:38,558,002-38,723,846	9p13.1		
			9p13	CNC	Vance et al. (1997)
CEN					
RP11-203L2	BES	chr9:70,447,920-70,642,602	9q21.11		
RP11-876N18	BES	chr9:70,831,740-71,036,759	9q21.11		

(continued)

Table 5.1 (continued)

		HSA9			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-63P12	AL135924	chr9:74,048,190-74,211,174	9q21.12		
RP11-522I20	AL354920	chr9:85,375,435-85,544,238	9q21.32		
RP11-30C23	AL451131	chr9:87,314,024-87,468,183	9q21.33		
RP11-507D14	AL137849	chr9:87,988,837-88,120,520	9q21.33		
RP11-155P1	BES	chr9:88,673,580-88,845,643	9q21.33		
RP11-107G16	BES	chr9:89,315,407-89,492,510	9q21.33		
RP11-164I22	BES	chr9:89,981,566-90,160,305	9q22.1		
RP11-875O18	BES	chr9:92,518,647-92,715,334	9q22.2		
RP11-714A6	BES	chr9:94,105,260-94,266,691	9q22.31		
RP11-240L7	BES	chr9:98,020,526-98,190,156	9q22.32		
RP11-330M2	AL158827	chr9:98,730,413-98,744,393	9q22.32		
RP11-106N7	BES	chr9:99,884,782-100,053,954	9q22.33		
RP11-208F1	BES	chr9:102,010,490-102,158,124	9q31.1		
RP11-354J3	BES	chr9:105,921,994-106,093,407	9q31.1		
RP11-714K8	BES	chr9:108,383,090-108,577,624	9q31.2		
RP11-18A3	AL359963	chr9:111,085,649-111,220,627	9q31.3		
RP11-243H16	BES	chr9:111,103,930-111,282,692	9q31.3		
RP11-16A3	BES	chr9:116,448,704-116,610,074	9q32		
RP11-336A17	AL160272	chr9:119,493,517-119,640,494	9q33.1		
		chr9:121.261.000-121,315,000		CNC(ChIP-on-chip)	Capozzi et al. (2008b)
RP11-100H1	BES	121.315	9q33.1		
RP11-160J24	BES	chr9:124,090,783-124,264,726	9q33.2		
RP11-542K23	AL359636	chr9:124,189,785-124,383,720	9q33.2		
				MMUENC	Ventura et al. (2004)
RP11-64P14	AL162254	chr9:124,304,812-124,493,132	9q33.2		
RP11-465F21	AC006313	chr9:124,622,045-124,630,661	9q33.2		
RP11-85O21	AC006450	chr9:125,657,313-125,834,867	9q33.3		

RP11-30A13	BES	chr9:131,628,209-131,798,961	9q34.11		
RP11-469E24	AL353636	chr9:137,865,306-137,936,843	9q34.3		

		HSA10			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-387K19	BES	chr10:149,098-312,071	10p15.3		
RP11-10D13	BES	chr10:214,415-366,376	10p15.3		
RP11-15D19	BES	chr10:835,011-1,011,342	10p15.3		
RP11-363N22	AL359878	chr10:854,871-1,039,159	10p15.3		
RP11-61P15	BES	chr10:13,747,461-13,911,792	10p13		
RP11-142F1	AL391334	chr10:17,555,784-17,653,214	10p12.3		
RP11-109I13	AL390783	chr10:18,510,777-18,688,434	10p12.3		
RP11-383B4	AL450384	chr10:18,842,308-18,966,878	10p12.4		
RP11-110M17	BES	chr10:24,276,209-24,449,403	10p12.1		
RP11-39E10	BES	chr10:31,183,319-31,363,437	10p11.23		
RP11-92J19	BES	chr10:36,759,042-36,945,342	10p11.21		
RP11-56L6	BES	chr10:38,038,398-38,212,095	10p11.21		
RP11-162G10	AL135791	chr10:38,123,110-38,190,084	10p11		
CEN					
RP11-351D16	AC010864	chr10:42,817,197-43,022,992	10q11.21		
RP11-285G1	AL353801	chr10:44,640,088-44,862,577	10q11.21		
RP11-90N8	BES	chr10:51,442,402-51,622,394	10q11.23		
RP11-1001A13	BES	chr10:52,032,752-52,229,644	10q11.23		
RP11-6J8	BES	chr10:57,715,995-57,887,658	10q21.1		
RP11-749A7	BES	chr10:63,201,531-63,375,079	10q21.2		
RP11-615M13	BES	chr10:78,282,179-78,448,176	10q22.3		
RP11-717O2	BES	chr10:84,795,601-84,982,857	10q23.1		
RP11-830J13	BES	chr10:88,357,707-88,550,596	10q23.2		
RP11-659F22	BES	chr10:89,246,763-89,428,545	10q23.3		
RP11-829M16	BES	chr10:92,526,680-92,680,990	10q23.31		

(continued)

Table 5.1 (continued)

		HSA10			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-166O7	BES	chr10:110,714,067-110,884,778	10q25.1		
		chr10:116,837,988-117,377,461	10q25.3	CNC(ChIP-ob-chip)	Lo et al. (2001a)
RP11-296H2	AL135793	chr10:123,907,460-124,121,322	10q26.13		
RP11-92A10	BES	chr10:132,010,830-132,165,851	10q26.3		
RP11-1022E21	BES	chr10:134,703,784-134,906,603	10q26.3		

		HSA11			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-401C19	AC083984	chr11:896,316-1,008,135	11p15.5		
RP11-650F7	BES	chr11:3,297,781-3,455,204	11p15.4		
RP11-749O23	BES	chr11:3,501,436-3,690,087	11p15.4		
RP11-661M13	BES	chr11:5,856,181-6,043,020	11p15.4		
				OWMENC	Cardone et al. (2007)
RP11-625D10	BES	chr11:5,667,339-5,864,725	11p15.4		
RP11-645I8	AC021935	chr11:6,072,745-6,229,122	11p15.4		
				PPY8ENC	Cardone et al. (2007)
RP11-56J22	BES	chr11:20,180,424-20,332,556	11p15.1		
RP11-103P20	BES	chr11:36,021,057-36,180,792	11p13		
RP11-150D18	BES	chr11:41,858,282-42,020,207	11p12		
			11p11	CNC	
RP11-29O22	BES	chr11:46,582,988-46,583,429	11p11.2		
RP11-318O24	BES	chr11:50,545,853-50,719,949	11p11.12		
CEN				GGOPTRHSAENC	Cardone et al. (2007)
RP11-217G11	BES	chr11:56,609,801-56,610,186	11q12.1		
RP11-75H24	BES	chr11:58,632,233-58,632,565	11q12.1		
RP11-160L9	BES	chr11:67,190,649-67,191,077	11q13.2		

RP11-955G14	BES	chr11:71,190,153-71,377,632	11q13.4		
RP11-757C15	AP000719	chr11:71,236,122-71,432,551	11q13.4		
RP11-807H22	AP000812	chr11:71,481,809-71,602,336	11q13.4		
RP11-7H7	BES	chr11:78,034,240-78,206,818	11q14		
RP11-119M23	BES	chr11:85,346,396-85,346,523	11q14.2		
RP11-529A4	AP004607	chr11:89,286,313-89,446,995			
			11q14.3	HLA11/NLE15ENC	Roberto et al. (2007)
RP11-692G6	BES	chr11:89,719,943-89,890,899			
RP11-732A21	AP001527	chr11:101,397,613-101,564,917	11q22.1		
RP11-864G5	AP000942	chr11:101,600,598-101,786,581	11q22.1		
RP11-1044B1	BES	chr11:105,109,962-105,322,691	11q22.3		
RP11-276O11	BES	chr11:105,262,409-105,262,775	11q22.3		
RP11-100J10	BES	chr11:112,570,375-112,735,819	11q23.1		
RP11-90A13	BES	chr11:130,889,654-131,037,422	11q25		
RP11-265F9	BES	chr11:134,272,267-134,441,179	11q25		
				APCEN	Cardone et al. (2007)

		HSA12			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-283I3	BES	chr12:153,051-329,683	12p13.33		
RP11-691J6	BES	chr12:5,200,006-5,384,129	12p13.32		
RP11-62G3	BES	chr12:6,121,261-6,298,431	12p13.31		
RP11-20D14	BES	chr12:8,690,273-8,864,148	12p13.31		
RP11-157L2	BES	chr12:9,788,001-9,945,600	12p13.31		
				NWMENC	Stanyon et al. (2008)
RP11-316E18	BES	chr12:9,916,001-10,122,368	12p13.31-p13.2		
RP11-13C13	BES	chr12:10,122,517-10,291,047	12p13.2		
RP11-502N13	BES	chr12:14,521,905-14,648,407	12p13.1		
RP11-1018J8	FISH	chr12:15,049,657-15,261,830	12p12.3		
RP11-489N6	FISH	chr12:16,084,282-16,171,229	12p12.3		

(continued)

Table 5.1 (continued)

		HSA12			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-871F6	FISH	chr12:17,423,813-17,640,601	12p12.3		
RP11-678N14	BES	chr12:19,566,006-19,721,709	12p12.3		
RP11-157I19	BES	chr12:20,050,190-20,206,352	12p12.2		
RP11-120A19	BES	chr12:20,325,423-20,486,337	12p12.2		
RP11-57F15	BES	chr12:20,863,387-21,018,324	12p12.2		
RP11-125O5	BES	chr12:21,151,769-21,303,642	12p12.2.p12.1		
RP11-12D15	BES	chr12:22,210,387-22,369,559	12p12.1		
RP11-877E17	BES	chr12:25,986,021-26,163,998	12p12.1		
RP11-666F17	FISH	chr12:26,671,081-26,857,010	12p11.23		
RP11-485K18	FISH	chr12:28,287,829-28,467,827	12p11.22		
RP11-517B23	BES	chr12:31,362,925-31,533,973	12p11.21		
RP11-956A19	BES	chr12:32,174,154-32,364,169	12p11.21		
RP11-460N10	FISH	chr12:33,170,516-33,333,493	12p11.1		
CEN					
RP11-152M7	BES	chr12:37,365,174-37,556,018	12q12		
RP11-490D11	BES	chr12:40,112,781-40,280,202	12q12		
RP11-618L22	AC079906	chr12:45,523,783-45,704,447	12q13.11		
RP11-23J18	BES	chr12:45,755,429-45,925,226	12q13.11		
RP11-241O10	BES	chr12:46,000,294-46,169,804	12q13.11		
RP11-47A12	BES	chr12:46,070,299-46,235,151	12q13.11		
RP11-19H5	BES	chr12:46,299,642-46,450,387	12q13.11		
RP11-254E3	BES	chr12:46,507,219-46,672,845	12q13.11		
RP11-30N17	BES	chr12:46,672,928-46,875,216	12q13.11		
RP11-159H4	BES	chr12:46,744,342-46,895,584	12q13.11		
RP11-204C20	BES	chr12:46,894,555-47,075,616	12q13.11		
RP11-94F1	BES	chr12:50,496,385-50,663,940	12q13.13		

RP11-699F3	BES	chr12:50,919,590-51,102,247	12q13.13		
RP11-4K11	BES	chr12:52,417,124-52,574,796	12q13.13		
RP11-631N16	BES	chr12:61,280,212-61,458,292	12q14.1-q14.2		
RP11-680F18	BES	chr12:63,441,896-63,614,208	12q14.3		
RP11-63J20	BES	chr12:80,424,584-80,582,696	12q21.31		
RP11-900F13	FISH	chr12:87,374,561-87,546,806	12q21.32-q21.33		
RP11-205I24	BES	chr12:102,490,342-102,647,694	12q23.3		
RP11-1G17	BES	chr12:110,393,571-110,596,386	12q24.12		
RP11-344G11	BES	chr12:125,063,337-125,209,987	12q24.32		
RP11-394D10	BES	chr12:132,034,089-132,208,159	12q24.33		

		HSA13			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
CEN					
RP11-110K18	AL137119	chr13:19,404,216-19,568,080	13q12.11		
RP11-45B20	AL445985	chr13:23,305,109-23,483,639	13q12.12		
				CMO18-CMO21ENCs	Cardone et al. (2006)
RP11-64I8	AL158065	chr13:30,406,381-30,571,172	13q12.3		
RP11-142E9	BES	chr13:33,252,754-33,451,136	13q13.2		
RP11-29G24	AL161718	chr13:34,851,469-34,910,004	13q13.3		
RP11-477C5	BES	chr13:41,599,503-41,760,120	13q14.11		
RP11-413N19	AL592523	chr13:41,969,072-41,973,065	13q14.11		
			LLA8ENC	Cardone et al. (2006)	
RP11-14553	BES	chrl13:45,340,029,42,504,601	13q14.11		
RP11-443J2	BES	chr13:45,279,141-45,450,817	13q14.12		
RP11-719B12	BES	chr13:45,408,175-45,579,860	13q14.12		
RP11-939G7	BES	chr13:45,754,269-45,939,953	13q14.13		
RP11-945G11	BES	chr13:45,928,366-46,127,167	13q14.13		
RP11-417C20	FISH	chr13:46,020,378-46,185,497	13q14.13		
RP11-103J18	AL138875	chr13:48,654,460-48,818,895	13q14.2		

(continued)

Table 5.1 (continued)

		HSA13			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-10O23	AC013618	chr13:55,430,704-55,602,978	13q21.1		
RP11-1043D14	BES	chr13:61,282,357-61,458,258	13q21.31		
				OWMENC	Cardone et al. (2006)
RP11-187E23	AL136999	chr13:66,092,979-66,264,337	13q21.32		
RP11-51P14	AL356006	chr13:67,146,127-67,174,887	13q21.32		
RP11-543G6	AC162212	chr13:70,669,808-70,794,225	13q21.33		
			13q21.33	CNC(ChIP-on-chip)	Cardone et al. (2006)
RP11-512J14	AL354995	chr13:70,797,636-70,947,217	13q21.33		
RP11-138N13	BES	chr13:74,311,795-74,458,502	13q22.2		
RP11-188A23	AL354831	chr13:77,153,157-77,297,742	13q22.3		
RP11-115N13	BES	chr13:82,035,688-82,200,947	13q31.1		
RP11-120L14	BES	chr13:83,766,483-83,924,544	13q31.1		
RP11-351H1	BES	chr13:84,396,772-84,582,561	13q31.1		
RP11-780G3	BES	chr13:85,161,451-85,333,456	13q31.1		
RP11-30L8	BES	chr13:85,529,117-85,655,956	13q31.1		
RP11-29P20	BES	chr13:86,954,636-87,112,529	13q31.2		
RP11-143O10	BES	chr13:88,496,254-88,673,921	13q31.2		
RP11-210E23	FISH	chr13:93,776,946-93,877,017	13q32.1		
RP11-721F14	BES	chr13:96,392,847-96,575,482	13q32.1		
RP11-46I10	BES	chr13:101,854,484-102,028,829	13q33.1		
RP11-261F2	AL445226	chr13:103,364,122-103,420,360	13q33.1		
RP11-245B11	AL161774	chr13:113,770,458-113,932,864	13q34		
RP11-569D9	FISH	chr13:113,930,807-114,103,243	13q34		
			13q34	CNC	Depinet et al. (1997)(case4)

		HSA14			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
CEN				HOMINOIDEA ENC	Ventura et al. (2003)
RP11-246M13(A)	BES	chr14:19,547,383-19,702,125	14q11.2		
RP11-68M15	BES	chr14:22,546,692-22,722,266	14q11.2		
RP11-3K11	BES	chr14:25,676,157-25,851,493	14q12		
RP11-96N22	BES	chr14:30,522,558-30,688,098	14q12		
RP11-642G19	BES	chr14:32,380,196-32,540,929	14q13.1		
RP11-918D6	BES	chr14:36,404,852-36,569,960	14q13.3		
RP11-94J22(B)	BES	chr14:41,623,645-41,782,055	14q21.1		
RP11-453F20	BES	chr14:44,679,792-44,872,979	14q21.3		
RP11-631K15	BES	chr14:48,752,711-48,915,809	14q22.1		
RP11-316E4	BES	chr14:50,001,799-50,183,814	14q22.1		
RP11-841O20	BES	chr14:52,073,343-52,285,417	14q22.1		
RP11-312M17	BES	chr14:54,251,694-54,407,050	14q22.2-.3		
RP11-81D11	BES	chr14:64,128,328-64,294,463	14q23.3		
RP11-886F16	BES	chr14:67,619,469-67,780,148	14q24.1		
RP11-204P19(C)	BES	chr14:71,001,855-71,164,272	14q24.2		
RP11-606A3	BES	chr14:73,138,310-73,312,477	14q24.2		
RP11-92H20	BES	chr14:74,381,660-74,551,240	14q24.2		
RP11-89I23	BES	chr14:79,486,870-79,652,196	14q31.1		
RP11-4E24	BES	chr14:85,025,843-85,182,785	14q31.3		
RP11-91C7	BES	chr14:90,549,220-90,692,115	14q32.12		
RP11-45E1	BES	chr14:96,885,041-97,054,197	14q32.2		
RP11-90G22	BES	chr14:100,210,924-100,389,009	14q32.2		
RP11-417P24	AL122127	chr14:105,267,349-105,437,150	14q32.33		
RP11-51P11(D)	BES	chr14:106,049,593-106,211,962	14q32.33		
				CMO13ENC	Ventura et al. (2003)

(continued)

Table 5.1 (continued)

		HSA15			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
CEN				HOMINOIDEA ENC	Ventura et al. (2003)
RP11-441B20(A)	AC080077	chr15:22,905,050-23,073,407	15q11.2		
RP11-570N16	AC019229	chr15:24,960,279-25,129,482	15q12		
RP11-11J16(B)	BES	chr15:29,518,297-29,691,905	15q13.3		
RP11-106G20(C)	BES	chr15:30,609,233-30,773,871	15q13.3		
RP11-50O2	BES	chr15:31,571,307-31,737,827	15q13.3		
RP11-747K21	BES	chr15:33,683,242-33,871,801	15q14		
RP11-720L8	AC068875	chr15:35,362,749-35,529,454	15q14		
RP11-133K1	AC020658	chr15:38,241,120-38,400,125	15q15.1		
RP11-729O24	BES	chr15:40,156,888-40,332,058	15q15.1		
RP11-753P14	BES	chr15:40,820,559-40,988,584	15q15.1		
RP11-594K13	BES	chr15:44,037,250-44,241,054	15q21.1		
RP11-846K6	BES	chr15:48,810,360-48,983,766	15q21.2		
RP11-316P21	AC025041	chr15:50,971,951-51,145,337	15q21.2		
RP11-126E3	BES	chr15:52,895,739-53,061,863	15q21.3		
RP11-450G20	BES	chr15:53,782,464-53,963,923	15q21.3		
RP11-294K12	BES	chr15:54,112,385-54,286,277	15q21.1		
RP11-844G16	BES	chr15:54,271,610-54,460,838	15q213		
RP11-829F13	BES	chr15:54,481,897-54,676,988	15q21.3		
RP11-323F24	BES	chr15:54,901,235-55,123,208	15q21.3		
RP11-44G18	ends	chr15:55,902,848-56,062,565	15q21.3		
RP11-93I17	BES	chr15:56,226,257-56,226,689	15q21.3		
RP11-236P11	AC087632	chr15:62,366,899-62,510,409	15q22.31		
RP11-282M16	AC022254	chr15:65,872,233-66,060,841	15q22.33		
RP11-1107A19(D)	ends	chr15:72,073,586-72,217,438	15q24.1		
			15q24.1	CNC	Ventura et al. (2003)
RP11-247C2(E)	AC010931	chr15:72,201,386-72,358,658	15q24.1		

RP11-624N5	ACO24552	chr15:72,158,366-72,251,969	15q24.1		
RP11-20M10	AC016276	chr15:75,965,634-76,127,696	15q24.2		
			15q25		Depinet et al. (1997)(case1)
RP11-1001M11(F)	BES	chr15:76,752,817-76,966,382	15q25.1		
				APCENNWMENC	Ventura et al. (2003)
RP11-16K12(G)	BES	chr15:76,939,472-77,105,720	14q25.1		
RP11-635O8	BES	chr15:80,103,012-80,257,524	15q25.2		
RP11-127F21	AC044907	chr15:81,186,929-81,348,689	15q25.2		
RP11-19E5(H)	AC027605	chr15:82,473,051-82,637,127	15q25.2		
			15q25.2	CNC	Ventura et al. (2003)
RP11-182J1(I)	AC048382	chr15:82,835,478-83,006,963	15q25.2		
			15q26.1	CNC	Rowe et al. (2000); Depinet et al. (1997)(case2)
RP11-90E5(J)	AC022710	chr15:98,163,252-98,349,768	15q26.3		

		HSA18			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-78H1	BES	chr18:2,136,811-2,307,213	18p11.32		
RP11-96I11	BES	chr18:12,904,782-12,904,961	18p11.21		
CEN					
RP11-10G8	BES	chr18:17,274,438-17,431,001	18q11.2		
RP11-104N11	BES	chr18:33,436,610-33,608,704	18q12.2		
RP11-61D1	AC090897	chr18:50,155,761-50,313,129	18q21.1		
				OWMENC	Ventura et al. (2007)
RP11-289E15	AC091135	chr18:50,360,135-50,526,341	18q21.2		
RP11-153B11	BES	chr18:52,818,203-52,977,905	18q21.2		
RP11-53N15	BES	chr18:70,195,436-70,195,693	18q22.3		
RP11-87C15	BES	chr18:75,965,206-75,965,502	18q23		

(continued)

Table 5.1 (continued)

		HSA20			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-371L19	AL118502	chr20:659,205-785,463	20p13		
RP5-1187M17	AL121891	chr20:3,013,541-3,139,396	20p13		
RP5-1068F16	AL023913	chr20:10,155,017-10,295,322	20p12.2		
		chr20:10,662,941-11,127,046	20p12.2	CNC(ChIP-on-chip)	Lo et al. (2001b)
RP4-813H11	AL079337	chr20:11,379,732-11,417,054	20p12.2		
RP5-1069O1	AL049633	chr20:15,126,843-15,219,199	20p12.1		
RP11-922G6	BES	chr20:22,887,406-23,046,035	20p11.21		
RP11-661H1	BES	chr20:23,454,246-23,627,064	20p11.21		
RP5-966J20	AL121925	chr20:24,698,120-24,737,379	20p11.21		
CEN					
RP11-1036L7	BES	chr20:28,048,230-28,206,006	20q11.1		
RP5-836N17	AL049539	chr20:30,126,905-30,238,598	20q11.21		
RP5-954P9	AL359828	chr20:34,046,335-34,084,879	20q11.23		
RP11-888D20	BES	chr20:34,932,840-35,111,176	20q11.23		
RP11-1152L20	BES	chr20:35,084,554-35,209,548	20q11.23		
RP11-192N1	BES	chr20:35,209,599-35,358,886	20q11.23		
RP11-826B14	BES	chr20:35,332,463-35,548,961	20q11.23		
RP11-138A15	BES	chr20:35,595,079-35,595,342	20q11.23		
RP5-906C1	AL133342	chr20:46,828,731-46,939,544	20q13.13		
RP5-1059L7	AL121913	chr20:55,665,561-55,815,784	20q13.32		
RP11-476I15	AL137028	chr20:62,376,540-62,435,964	20q13.33		
				CMO22ENC	Misceo et al. (2005)

		HSA20			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP11-800K15	BES	chrX:483,105-664,235	Xp22.33		
		chrX:6,000,001-9,500,000	Xp22.31	CNC	
RP11-458E23	BES	chrX:10,007,515-10,251,587	Xp22.2		
RP11-450P7	AL772392	chrX:21,383,521-21,507,706	Xp22.12		
RP11-450E21	AL591591	chrX:33,274,317-33,378,433	Xp21.1		
RP11-64P15	BES	chrX:33,376,434-33,542,431	Xp21.1		
RP11-1078G21	BES	chrX:33,512,076-33,704,078	Xp21.1		
RP11-825L2	BES	chrX:33,920,685-34,107,259	Xp21.1		
RP11-281B1	BES	chrX:33,989,930-34,174,295	Xp21.1		
RP11-910L4	BES	chrX:34,033,952-34,208,856	Xp21.1		
RP11-831J15	BES	chrX:34,148,053-34,301,011	Xp21.1		
RP11-384A17	BES	chrX:43,240,049-43,392,966	Xp11.3		
RP11-552J9	AL450023	chrX:52,556,131-52,566,971	Xp11.22		
CEN					
RP11-978L24	BES	chrX:61,470,646-61,691,665	Xq11.1		
RP11-148E15	BES	chrX:62,253,894-62,418,154	Xq11.1		
RP11-135B16	BES	chrX:62,460,317-62,628,230	Xq11.1		
RP11-213M6	BES	chrX:62,791,311-62,954,364	Xq11.1		
RP11-151C15	BES	chrX:62,874,379-63,050,505	Xq11.1		
RP11-754F6	BES	chrX:63,033,136-63,192,316	Xq11.1		
RP11-346J4	BES	chrX:63,171,662-63,365,730	Xq11.1		
		chrX:65,100,001-67,700,000	Xq12	CNC	
RP11-625B4	BES	chrX:69,721,202-69,884,160	Xq13.1		
RP11-395L12	AL157933	chrX:81,134,235-81,183,002	Xq21.1		
RP11-483J19	BES	chrX:92,542,566-92,694,921	Xq21.32		
			Xq21.2	LCAXENC	Ventura et al. (2001)
RP11-449F11	FISH	chrX:96,896,057-97,059,042	Xq21.33		
RP11-426L6	BES	chrX:104,850,550-105,005,976	Xq22.3		

(continued)

Table 5.1 (continued)

		HSA20			
BAC name	Acc.N.	UCSCMarch2006	Cytog.map	ENC or CNC	Reference
RP5-874H6	AL078580	chrX:111,900,664-111,922,369	Xq23		
RP11-243N2	BES	chrX:115,064,564-115,228,958	Xq23		
RP11-488B15	BES	chrX:124,895,774-125,047,349	Xq25		
RP11-535K18	AL078638	chrX:134,948,985-135,131,392	Xq26.3		
RP11-478P19	BES	chrX:143,327,218-143,502,603	Xq27.3		
RP11-402H20	AC016977	chrX:153,772,076-153,951,934	Xq28		

The table reports, for each chromosome, a panel of BAC clones used to delineate its evolutionary history in primates, essentially as reported by Stanyon et al. (2008), in the Supplementary files. Chromosomes not showing any ENC or finely-mapped human clinical centromeres are not reported. For these chromosomes the reader can refer to Fig. 5.1 and to Marshall et al. (2008). The first column shows the BAC name; a letter in parenthesis after the BAC name, occasionally reported, indicates the BAC code utilized in Figs. 5.3 and 5.5. The second column indicates the method used for placing the BAC on the human sequence (BES = BAC End Sequence; see Sect. 5.6), reported in the third column, while its cytogenetic position is shown in column four.

In this frame, the table reports, in the fifth column:

1. The ENCs (hatched red row) and the corresponding reference (sixth column). Usually, reiterative FISH experiments have been performed to characterize at the maximal resolution the mapping of each ENC. The closest BACs on each side of the ENC are reported. The acronyms of the species in which the ENC has been discovered are reported below.
2. The clinical neocentromeres (hatched light-blue rows) that have been mapped at least at a cytogenetic band resolution. The annotation "ChIP" indicates that they have been mapped by ChIP-on-chip technology (see text). In this case the CENP-A or -C domains is reported.
3. The three human repositioned centromeres (HRC)
4. The normal human centromere (blue rows)

ENC Evolutionary new centromere, *CNC* Clinical neoCentromere, *HRC* Human repositioned centromere, *AC* Ancestral centromere, *AP* Ancestral primate
Literature not reported in the main paper is reported below.
Species' acronyms: CJA Callithrix jacchus (common marmoset) (NWM), *CMO Callicebus moloch,* also indicated as *Callicebus pallescens* (dusky titi) (NWM), *GGO Gorilla gorilla* (gorilla), *HLA Hylobates lar* (lar gibbon), *LCA Lemur catta* (ring-tailed lemur), *MMU Macaca mulatta* (rhesus monkey), *PPY Pongo pygmaeus* (orangutan)

Chapter 6
Structure and Evolution of Plant Centromeres

Kiyotaka Nagaki, Jason Walling, Cory Hirsch, Jiming Jiang, and Minoru Murata

Contents

K. Nagaki and M. Murata (✉)
Research Institute for Bioresources, Okayama University, Kurashiki 710-0046, Japan
e-mail: mmura@rib.okayama-u.ac.jp

J. Walling, C. Hirsch, and J. Jiang
Department of Horticulture, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA
e-mail: jjiang1@wisc.edu

Đ. Ugarković (ed.), *Centromere,* Progress in Molecular and Subcellular Biology 48,
DOI: 10.1007/978-3-642-00182-6_6, © Springer-Verlag Berlin Heidelbarg 2009

Abstract Investigations of centromeric DNA and proteins and centromere structures in plants have lagged behind those conducted with yeasts and animals; however, many attractive results have been obtained from plants during this decade. In particular, intensive investigations have been conducted in *Arabidopsis* and Gramineae species. We will review our understanding of centromeric components, centromere structures, and the evolution of these attributes of centromeres among plants using data mainly from *Arabidopsis* and Gramineae species.

6.1 Introduction

The centromere is a functional chromosomal site that helps to divide sister chromatids equally into daughter cells in mitotic and meiotic cell divisions. Important functions include cohesion and separation of sister chromatids, attachment of spindle fibers, chromosomal segregation, and the control of cell-cycle checkpoints. The centromere is integral for the control of these functions. Usually, one centromere is formed on a chromosome at a primary constriction site. At metaphase, spindle fibers attach and pull sister chromatids towards different poles to divide the chromatids into daughter cells. A complex of centromeric DNA and proteins is formed at the primary constriction, and this complex is called a kinetochore. In the following chapters, we will review centromeric components, centromere structures, and the evolution of these attributes among plants.

6.2 Centromeric DNA

6.2.1 General Remarks

Although centromeres have a highly conserved role of transmitting chromosomes to subsequent generations, among eukaryotes, centromere DNA is highly variable among species. For example, only a 125-bp DNA sequence is necessary for centromere function in budding yeast, *Saccharomyces cerevisiae* (Cottarel et al. 1989). A single centromere-specific nucleosome is formed at the 125-bp DNA sequence (Furuyama and Biggins 2007), and the nucleosome recruits other centromeric proteins to construct a kinetochore (Amor et al. 2004). In contrast to budding yeast, higher-level eukaryotes have more complex centromeric DNA. For example, the fission yeast *Schizosaccharomyces pombe* possesses a 30-kb centromeric DNA sequence that forms centromeric nucleosomes (Choo 1997). Yet, the centromeric DNA of budding yeast has no sequence similarity with the centromeric DNA of the fission yeast. The centromeric DNA of multicellular eukaryotes is even more complex than that of yeast. The centromeres of multicellular eukaryotes usually consist of tandem repetitive DNA sequences, and the size of the centromere can be of several mega

bases in length. These repeat arrays in humans are called alpha (α) satellites and are composed of a basal 171-bp repeat arranged in tandem arrays that range in size from 250 kb to 4 Mb, with each centromere harboring different amounts (Wevrick and Willard 1989). As with human centromeres and those of other mammals, plant centromeres also have mega-base sized arrays of tandem repetitive DNA sequences. Additionally, transposable elements are also abundant in centromeric and paracentromeric regions. Reported sequences in plant centromeres are listed in Table 6.1.

6.2.2 Arabidopsis

A clone containing a 180-bp repeat family, pAL1, was cloned from *Arabidopsis thaliana* (Martinez-Zapater et al. 1986) and its chromosomal localization was checked using fluorescence in situ hybridization, FISH (Maluszynsak and Heslop-Harrison 1991; Murata et al. 1994). The FISH signals were observed on centromeric regions of all five *A. thaliana* chromosomes. Although other centromere-specific and nonspecific repetitive sequences including *Athila* were also found on centromeric regions in *A. thaliana* (The Arabidopsis genome initiative 2000), only 180-bp family sequences were co-precipitated using chromatin immunoprecipitation (ChIP) with anti-HTR12 (a centromere-specific histone H3 in *A. thaliana*, described in Sect. 6.3.2) antibody (Nagaki et al. 2003). In a close relative of *A. thaliana*, *A. arenosa*, a tandem repetitive DNA family of ca. 170-bp repetitive units, pAa, was isolated and subsequent analyses showed that the sequence is located on all 32 centromeres of the species (Kamm et al. 1995). The pAa sequences share 50–80% sequence similarity with pAL1 sequences. The sequence was also observed on 16 of the 26 chromosomes of *A. suecica*, while the 180-bp family sequence is observed on the rest of the chromosomes (10 of the 26 chromosomes). This implies *A. suecica* is a hybrid species of *A. thaliana* and *A. arenosa*, and both of the two different centromeric DNA sequences are retained in the hybrid species. Additionally, *A. pumila* and *A. griffithiana* also have a species-specific subfamily of a 180-bp family sequence (Heslop-Harrison et al. 2003). The existence of these species-specific subfamilies of 180-bp family sequences suggests that ancestral 180-bp family sequences have diverged in descendant species, and these sequences are an established centromeric-DNA component of all of the *Arabidopsis* species. Although most diploid species only retain a single centromeric tandem repeat, exceptional examples were found in *A. halleri* and *A. lyrata* (Kawabe and Nasuda 2005). These species are closely related to *A. arenosa* and possess pAa sequences. However, in addition to pAa sequences, these species also have two species-specific 180-bp repeat subfamilies, pAge1 and pAge2. Four of the eight *A. halleri* centromeres possess pAa, one of the eight possesses pAge1, two of the eight have pAge2, and the remaining centromere possesses pAs and pAge1. Since repetitive DNA is thought to adapt to its associated centromeric proteins and therefore is selected for by the proteins from a repetitive DNA sequences pool (Dawe and Henikoff 2006), its possible that these particular

Table 6.1 Known plant centromeric DNA sequences

Species	Repeat	Size (bp)	Type	ChIP	Refs
Arabidopsis arenosa	pAa	166–179	T	NT	Kamm et al. (1995)
Arabidopsis gemmifera	pAge1	180	T	NT	Kawabe and Nasuda (2005)
	pAge2	180	T	NT	Kawabe and Nasuda (2005)
Arabidopsis griffithiana	pAgKB1	180	T	NT	Heslop-Harrison et al. (2003)
Arabidopsis pumila	pApKB2	180	T	NT	Heslop-Harrison et al. (2003)
Arabidopsis thaliana	180-bp repeat family	180	T	P	Martinez-Zapater et al. (1986); Murata et al. (1994); Nagaki et al. (2003)
	Athila	10,500	R	NP	Nagaki et al. (2003); Pelissier et al. (1995); The Arabidopsis genome initiative (2000)
Beta corolliflora	pHC8	162	T	NT	Gindullis et al. (2001b)
Beta procumbens	pTS5	158–160	T	NT	Schmidt and Heslop-Harrison (1996)
	pTS4.1	312	T	NT	Schmidt and Heslop-Harrison (1996)
	pBp10	417	R	NT	Gindullis et al. (2001b)
Beta vulgaris	pBV1	326–327	T	NT	Schmidt and Metzlaff (1991)
	pBv26	417	R	NT	Gindullis et al. (2001b)
Brachycome dichromosomatica	Bd49	176	T	NT	Leach et al. (1995)
Brachypodium sylvaticum	CCS1	480	R	NT	Aragon-Alcaide et al. (1996)
Brassica campestris	pBcKB4	175	T	NT	Harrison and Heslop-Harrison (1995)
Brassica oleracea	pBoKB1	171	T	NT	Harrison and Heslop-Harrison (1995)
Hordeum vulgare	(AGGGAG)n satellite	6	T	P	Houben et al. (2007); Hudakova et al. (2001)
	cereba (CR family)	7,176	R	P	Houben et al. (2007); Hudakova et al. (2001); Presting et al. (1998)
Oryza brachyantha	CentO-F	154	T	P	Lee et al. (2005)
Oryza rhizomatis	CentO-C1	126	T	P	Lee et al. (2005)
	CentO-C2	366	T	P	Lee et al. (2005)
Oryza sativa	CentO	155	T	P	Cheng et al. (2002); Dong et al. (1998); Nagaki et al. (2004); Nonomura and Kurata (1999)

	CRR (CR family)	7,400–7,800	R	P	Cheng et al. (2002); Dong et al. (1998); Nagaki et al. (2004, 2005b); Nonomura and Kurata (1999)
Pennisetum glaucum	pPgKB19	137	T	NT	Kamm et al. (1994)
Petunia hybrida	pBS-SB1-B5	666	T	NT	Entani et al. (1999)
Pinus densiflora	PDCD501	27	T	NT	Hizume et al. (2001)
Saccharum officinarum	SCEN	140	T	P	Nagaki and Murata (2005); Nagaki et al. (1998)
	CRS (CR family)		R	P	Nagaki and Murata (2005)
Secale cereale	Bilby	3,400	R	NT	Francki (2001)
Solanum bulbocastanum	pSbTC1	7	T	NT	Tek and Jiang (2004)
Solanum lycopersicum	TGRIV	7,000	R	NT	Chang et al. (2008)
Sorghum bicolor	pSau3A10	137	T	NT	Miller et al. (1998)
	pSau3A9 (CR family)		R	NT	Jiang et al. (1996); Miller et al. (1998)
Torenia bailonii	BCEN-family	52	T	NT	Kikuchi et al. (2005)
Torenia fournieri	TCEN-family	52	T	NT	Kikuchi et al. (2005)
Triticum aestivum	TaiI	570	T	NT	Kishii et al. (2001)
	pBS301	250	T	NT	Cheng and Murata (2003)
	CRW (CR family)	7,762–7,865	R	P	Liu et al. (2008)
Vigna unguiculata	pVuKB1	488	T	NT	Goel et al. (2002)
Zea mays	CentC	156	T	P	Ananiev et al. (1998); Zhong et al. (2002)
	Cent4	740	T	NT	Page et al. (2001)
	CRM (CR family)	7,572	R	P	Ananiev et al. (1998); Zhong et al. (2002)
	B repeat	540	T	NT	Alfenito and Birchler (1993)
Zingeria biebersteiniana	Zbcen1	755	T	NT	Saunders and Houben (2001)
	Zb47A		R	NT	Saunders and Houben (2001)

Abbreviations: *T* tandem repetitive sequence, *R* retrotransposon related sequence, *NT* not tested, *P* precipitated with CENH3 antibodies by ChIP, *NP* not precipitated with CENH3 antibodies by ChIP

species might be in the middle of the selection process. On the other hand, that these species possess two HTR12 genes suggests that these HTR12 proteins may be driving the selection independently (Kawabe et al. 2006).

6.2.3 Graminaceae

The centromeres of several species within the family Graminaceae represent the largest group of closely related plant species in which the DNA composition of centromeres has been extensively studied (Table 6.1). Despite the reported high degree of colinearity among most grass genomes (Gale and Devos 1998), the sequences of their centromeres are, surprisingly, quite variable in terms of centromere size, repeat abundance, and arrangement of repeats. More recent studies have provided evidence that allow these repeats to be arranged into two groups: centromeric satellites and centromere-specific retrotransposons (CRs) (Cheng et al. 2002; Zhong et al. 2002; Nagaki et al. 2005b).

Centromeric satellites within the family Graminaceae have been reported for several species, including rice (CentO), maize (CentC), sugarcane (SCEN), sorghum (pSau3A10), and barley (GC-rich microsatellite) (see Table 6.1). Among the cereals, CentO and CentC have the most extensively studied satellite repeats. In cultivated rice (*Oryza sativa* cv. Nipponbare), CentO monomers are 155 bp in length, located on each of the twelve chromosomes, and range in total array size from 65 kb to 2 Mb (Cheng et al. 2002). The abundance of CentO in rice varies even between the two subspecies, with reports of *japonica* varieties containing five times less CentO than the homologous centromere region in the *indica* variety (Cheng et al. 2002). When compared to each other, the sequences of CentO and CentC display relatively short domains of similarity (Cheng et al. 2002). However, despite this observation, most satellite repeats show surprisingly little homology across species and are therefore generally considered species specific (Henikoff et al. 2001). Furthermore, a wild species of rice (*O. brachyantha*) that diverged from cultivated rice less than ten million years ago (Ge et al. 1999) has completely lost CentO and replaced it with a novel satellite array that shows no sequence homology to repeats in other species within the genus *Oryza* (Lee et al. 2005). These findings suggest that this dominant and highly represented component of cereal centromeres can undergo rapid evolutionary changes and is extraordinarily dynamic at the sequence level.

The centromeres of Graminaceae species contain a distinct centromere-specific retrotransposon family (CR family). Sequences related to the CR elements were first reported in *Brachypodium* (CCS1) and sorghum (pSau3A9) and have been found in all grasses interrogated for such sequences, including rice (CRR), maize (CRM), wheat (CRW), barley (*cereba*), and sugarcane (CRS) (see Table 6.1). Most CRs belong to the *Ty3-gypsy* family of retrotransposons, with their protein-coding domains flanked by long terminal repeats (LTRs) on each side. The myriad of intact and solo LTR retrotransposons in centromere regions, and the lack of orthology

among anchored LTRs in related species, suggest that CRs are a dynamic component of the centromere and are continually being reorganized (Ma and Jackson 2006).

Like the satellite repeats, FISH and co-immunoprecipitation experiments have confirmed that CRs are bona fide constituents of functional centromeres. CRs typically have a larger distribution across the centromere than the satellite array and are also found in the pericentromeric region (Nagaki et al. 2005b). Although tandem arrays of CRs have been reported, their arrangements within the centromere appears more sporadic than the satellites and they are often found inserted internally within the larger satellite arrays and even nested within themselves (Cheng et al. 2002; Jin et al. 2004). The degree of CR intermingling with satellites is variable. For example, FISH on extended DNA fibers (fiber FISH) analysis using CRM and CentC in maize has been used to measure tracks of centromere-specific satellites of over 2 Mb in length in which CRM is extensively intermingled throughout the array (Jin et al. 2004). In rice, however, similar approaches have led to the conclusion that CRR intermingling over the length of CentO is more irregular and often interrupted, with stretches of CentO up to 400 kb in length that are devoid of any CRR (Cheng et al. 2002). In wheat, the centromere region contains arrays of repetitive DNA arranged in intervals of up to 55 kb (Fukui et al. 2001) that all seem to have evolved directly from CRW elements, suggesting that the maintenance and growth of the arrays result from the amplification and reshuffling of the basal retroelement (Liu et al. 2008).

6.3 Centromeric Proteins

6.3.1 General Remarks

Despite the exceptional degree of variability of centromeric DNA among species, many centromeric proteins are highly conserved (Amor et al. 2004). Centromeric proteins have been intensively investigated in both yeast and mammals and these studies have resulted in the characterization of several centromere-specific proteins. In the fission yeast, centromeric proteins were first identified from mini chromosome instability (Mis) mutants (Takahashi et al. 1994). In humans, the centromeric proteins CENP-A, -B, and –C were first identified as antigens from autoimmune disease patients (CREST) (Earnshaw and Rothfield 1985). Recently, immunoprecipitated human kinetochore complexes were investigated using MS spectroscopic analysis, which found more than 40 centromeric proteins that were included in the complexes (Obuse et al. 2004; Okada et al. 2006).

At the beginning of plant centromeric protein investigations, the antisera of CREST patients were tested to determine whether they could cross-react with plant centromeric proteins. Results indicated that a few of the sera recognized centromeric regions of plants, implying that at least a portion of the centromeric proteins is shared among plants and animals (Mole-Bajer et al. 1990; Houben et al. 1995). In plants, the first centromeric protein that was confirmed to be localized to the

centromere region was CENP-C homologs in maize (Dawe et al. 1999). Subsequent investigations of plant centromeric proteins have led to the identification of CENP-A and Mis12 homologs from *A. thaliana* (Talbert et al. 2002; Sato et al. 2005), and homologs of these proteins were subsequently identified from other plants (Ogura et al. 2004; Zhong et al. 2002; Nagaki et al. 2004, 2005a; Nagaki and Murata 2005). Other plant centromeric proteins have also been reported, and these plant centromeric proteins are listed in Table 6.2.

6.3.2 CENH3

The centromere can be described at the molecular level by the replacement of canonical histone H3 with a specialized centromere histone H3 variant, CENH3. The first CENH3 discovered was CENP-A of humans (Palmer et al. 1987, 1991), and since its discovery CENH3 genes have been found in all eukaryotes researched, including yeast (Cse4) (Meluh et al. 1998), *Drosophila melanogaster* (CID) (Henikoff et al. 2000), *A. thaliana* (HTR12) (Talbert et al. 2002), and rice (CENH3) (Nagaki et al. 2004). The CENH3 protein, like canonical histone H3, has two domains: a N-terminal tail domain and a histone fold domain (HFD). Sequence homology in the HFD is seen between CENH3 and H3 both within and between species. CENH3 can be distinguished from the more abundant histone H3 by its N-terminal tail domain, which is not similar in its DNA sequence or base-pair length to the H3 histone both within or between species (Malik and Henikoff 2001).

CENH3 replaces H3 on active centromeric DNA interacting with other histone proteins and is necessary for the proper formation of the kinetochore (Choo 2001; Henikoff et al. 2001). The use of FISH and ChIP has led to the discovery that CENH3 binds to the centromere repeats CentO/CRR and CentC/CRM of rice and maize, respectively. Furthermore, CENH3 does not associate with all of the CentO/CRR or CentC/CRM repeats (Jin et al. 2004; Nagaki et al. 2004). For instance, the centromeres of numerous species, including rice, do not contain a continuous string of CENH3 nucleosomes, but rather an intermingling of CENH3 with canonical H3 (Blower et al. 2002; Nagaki et al. 2004). This is further explained through models in which CENH3 nucleosomes are constrained to the outer regions of the chromatid that interacts with the microtubules, while H3 containing heterochromatic nucleosomes are restricted to the inner regions that promote sister chromatid cohesion (Blower et al. 2002).

The CENH3 proteins found in different species are functionally conserved. Typically, when a gene is conserved in function, it is also fairly conserved at the sequence level as well, but this is not seen for CENH3 genes (Malik and Henikoff 2001). In addition, the centromere DNA repeats with which CENH3 interacts are also highly diverged (see Sect. 6.2). This realization led to the finding that suggests CENH3 of *D. melanogaster* (CID) is adaptively evolving (Malik and Henikoff 2001; see Chap. 2 in this book). Adaptive evolution of CENH3 has also been proposed in plants. Talbert et al. (2002) compared the CENH3 gene (HTR12) from *A. thaliana* and *A. arenosa* and found evidence for adaptive evolution in the

Table 6.2 Known plant centromeric proteins

		Plant		
Type	Mammal	Crossreacted with anti-non-plant centromeric protein antibody	DNA sequence was found	Checked by original antibodies
Assembly	CENP-A (CENH3)		SoCENH3 (Nagaki and Murata 2005)	HTR12 (Talbert et al. 2002), Maize CENH3 (Zhong et al. 2002), OsCENH3 (Nagaki et al. 2004)
	CENP-C		Barrel medic CENP-C, Potato CENP-C, Tomato CENP-C, Beet CENP-C, and Black cottonwood CENP-C (Talbert et al. 2004)	Maize CENP-C (Dawe et al. 1999), AtCENP-C (Ogura et al. 2004)
	CENP-F	CENP-F (ten Hoopen et al. 2000)		
	Dyskelin	CBF5 (ten Hoopen et al. 2000)		
	hMis12			AtMis12 (Sato et al. 2005)
	Meiotic histone			*Lilium longiflorum* MH (Suzuki et al. 1997)
	NDC80			Maize NDC80 (Du and Dawe 2007)
	p19Skp1	SKP1 (ten Hoopen et al. 2000)		
Movement	CENP-E			Cpel1 and Cpel2 (ten Hoopen et al. 2002)
Checkpoint	3F3/2	3F3/2 (Yu et al. 1999)		
	hBub1		Bub1-like (Houben and Schubert 2003)	
	Bub3		Bub3-like (Houben and Schubert 2003)	
	hMad2			Maize Mad2 (Yu et al. 1999)
	hZw10		Zw10 (Starr et al. 1997)	

N-terminal tail of the protein. This analysis was extended to include more members of the Brassicaceae and revealed not only adaptive evolution in the N-terminal tail, but also adaptive evolution in the more conserved HFD, including the loop 1 region of the HFD (Cooper and Henikoff 2004). The loop 1 region of CENH3 is important because it is necessary and sufficient for CENH3 localization to centromeres (Vermaak et al. 2002). These findings lead to the arms race hypothesis in which centromere DNA repeats are changing and expanding to increase their segregation properties, while CENH3 is changing to curb this and keep segregation frequencies equal to avoid fixing traits (Malik and Henikoff 2001; Talbert et al. 2002; see Chap. 2 in this book).

6.3.3 CENP-C

CENP-C is one of the centromeric proteins isolated as an antigen from CREST patients. CENP-C exhibits DNA-binding properties and is located at the inner kinetochore plate in humans (Saitoh et al. 1992; Yang et al. 1996). Disruptions of CENP-C homologs have resulted in mitotic delay and abnormality regarding chromosome segregations in vertebrates (Fukagawa and Brown 1997; Kalitsis et al. 1998). Homologs of human CENP-C have been isolated from various eukaryotes, including yeasts, animals, and plants, and a comparative analysis divided these homologs into three kingdom-consented subfamilies (Dawe et al. 1999; Ogura et al. 2004; Talbert et al. 2004). The conserved sizes of the subfamilies were ca. 940 amino acids (aa) in animals, ca. 550 aa in yeasts, and ca. 700 aa in plants, and yet only a 24-aa motif, the CENP-C motif, was conserved among these sequences (Talbert et al. 2004). Although C-terminal regions of plant CENP-C homologs including the CENP-C motif are highly conserved, N-terminal regions show limited sequence similarity among plant CENP-C homologs. Furthermore, two pairs of exons in the middle region of grass species have been duplicated, deleted, and positively selected during their evolution (Talbert et al. 2004). In addition to data from grass species, a comparative analysis between CENP-C of *A. thaliana* and *A. arenosa* uncovered adaptive evolution of the N-terminal regions of CENP-C among plants (Talbert et al. 2004). Cytological localizations of plant CENP-C homologs were investigated in maize and *A. thaliana* by immunostaining using species-specific anti-CENP-C antibodies, and results showed the continuous existence of the CENP-C homologs on their centromeres throughout their cell cycles (Dawe et al. 1999; Ogura et al. 2004).

6.3.4 Mis12

Mis12, first isolated from fission yeast, was identified as one of constitutive centromeric proteins, and mutants of this protein were shown to induce the unequal segregation of chromosomes (Goshima et al. 1999). The human homolog (hMis12)

also exhibited centromeric localization in human cells, and RNA interference of hMis12 induces chromosome misalignment and missegregation in human cells (Goshima et al. 2003). Mis12 homologs were also found in two plant species: *A. thaliana* and *Glycine max* (soybean) (Goshima et al. 2003). Additionally, Mis12 homologs were surveyed for other plant species, and three additional homologs were found in rice, bread-wheat, and grape (Sato et al. 2005). These Mis12 homologs possess similar sizes (259 aa in *S. pombe*, 205 aa in humans, and 238–249 aa in plants) and two conserved blocks at the N-terminal regions (Goshima et al. 2003; Sato et al. 2005). Chromosomal localization of the *Arabidopsis* homolog (AtMIS12) showed co-localization with HTR12 on *Arabidopsis* centromeres (Sato et al. 2005). Although the location of AtMIS12 overlapped with that of HTR12 in almost all regions, AtMIS12 occupied only a part of the 180-bp repeat family sequence tracts (Sato et al. 2005).

6.4 Structure of Plant Centromeres

6.4.1 Arabidopsis

The genetic positions of all five *Arabidopsis* centromeres were determined using a mutant that produces a nonseparated tetrad of pollen grains, *qrt1* (Preuss et al. 1994; Copenhaver et al. 1999). In the *Arabidopsis* genome sequencing project, the components and structures of the five centromeres were partially uncovered, but large gaps remain in the middle of all centromeres (The Arabidopsis genome initiative 2000). A total of 5 Mb of partial DNA sequences from the five centromeres were indentified and analyses using the sequences revealed that the DNA consisted of various kinds of repetitive DNA sequences including transposons, retrotransposons, microsatellites, and tandem repeats (The Arabidopsis genome initiative 2000). A total of 47 expressed genes were found in the pericentromeric regions. To uncover sequences in the middle of the centromeres, a physical map was constructed using DNA from a hypomethylated strain, *ddm1*. Genome walking using BAC libraries followed by sequencing of the tiled BAC clones were used to reveal the fine structure of this region (Kumekawa et al. 2000, 2001; Hosouchi et al. 2002). In the centromeric region of chromosome 5, 180-bp repeat family sequences are tandemly repeated at both edges of the central domains, but the orientations of the repeat tracts are inverted (Kumekawa et al. 2000). Various kinds of transposable elements were inserted into the flanking regions of the centromeric region of chromosome 5, while the central domain preferentially accumulated the element: *Athila*. The sizes of the genetically mapped centromere and the central domain of chromosome 5 were determined to be 4.7 and 2.9 Mb, respectively (Kumekawa et al. 2000; Hosouchi et al. 2002). Although the centromeric region of chromosome 4 also showed similarities to the insertions patterns of the transposable element within the centromeric region of chromosome 5, the inverted positioning of the

180-bp repeat family tracts were not observed in the centromeric region of chromosome 4 (Kumekawa et al. 2001). The sizes of the genetically mapped centromere and the central domain of chromosome 4 were determined to be 5.3 and 2.7 Mb, respectively (Kumekawa et al. 2001). Additionally, the sizes of the genetically mapped centromeres of chromosome 1, 2, and 3 were determined as 9, 4, and 4 Mb, respectively (Hosouchi et al. 2002). Data from ChIP using anti-HTR12 antibody and immunostaining on extended chromosome indicated that not all of the 180-bp repeat family sequences in the central domains were co-localized with HTR12 (Nagaki et al. 2003; Shibata and Murata 2004). The data suggest that a part of the core domain is acting as a functional centromere.

6.4.2 Structure and Evolution of Centromere 8 in Rice

Despite the increasing support of robust genome-wide sequencing data, determining the molecular structure of centromeres in higher eukaryotes has evaded researchers. The abundance of satellite repeats in centromeres has largely precluded any efforts to fully sequence centromeres in higher eukaryotes. In the grasses, data are slowly emerging that shed light on these enigmatic areas of the genome. Current reports of centromere structure within the Graminaceae are for the most part restrained to characterizations of individual sequence components of centromeres (see Sect. 6.2.3 and Table 6.1). Although these findings provide a foundation for centromere research and illuminate some key elements, they still leave much to be determined in terms of revealing the overall structure and dynamics of a functional centromere. It was recently discovered in cultivated rice that the centromere of chromosome 8 (*Cen8*) does not contain the abundance of satellite DNA that is usually present in most centromeres (Cheng et al. 2002). The paucity of satellites harbored in this centromere is a key characteristic that allowed this centromere to be fully sequenced and, as such, has bolstered the use of rice as a leading model for studying centromere structure and evolution (Nagaki et al. 2004).

At the cytological level, *Cen8* resembles the other rice centromeres in terms of size and position (Cheng et al. 2002). Genetically, this centromere is defined as a region of little or no detectable recombination (<1 per 186 recombinant events) and is positioned at approximately 54.0 cM on a linkage map derived from an *O. sativa* ssp. japonica/*O. sativa* ssp. indica cross (Harushima et al. 1998). This recombination-free zone spans approximately 2.3 Mb of DNA and contains two distinct centromere-specific sequence elements, CentO satellite DNA and CRR; it also contains one enveloping domain of chromatin that is enriched with CENH3. These three components collectively define the centromere region on chromosome 8 in rice (Nagaki et al. 2004; Wu et al. 2004; Yan et al. 2005).

The CENH3-binding domain of *Cen8* is approximately 750 kb as revealed by the use of ChIP with antibodies against CENH3 (Nagaki et al. 2004). However, the centromere chromatin in the core domain is not exclusively composed of CENH3 histone, but rather alternating tracts of canonical histones and CENH3 histones.

It has been suggested that the mingling of canonical histones with centromere histones might provide a basis for the assembly of higher-level structures that effectively limit kinetochore/microtubule binding to only the outside of chromatids (Sullivan and Karpen 2004; Black and Bassett 2008).

Repetitive DNA has long been known to dominate in the centromeres of eukaryotes. *Cen8* is no exception, with satellite arrays and centromere retrotransposons collectively accounting for more than 60% of the centromere region analyzed (Nagaki et al. 2004; Wu et al. 2004). However, the representation of satellite DNA (CentO) in *Cen8* is low compared with most of the other rice centromeres. *Cen8* contains approximately 65 kb of satellite DNA nested within the recombination-free region, where a few other centromeres of rice such as *Cen1* and *Cen11* are estimated to contain over a mega-base of this array within a similar-sized region (Cheng et al. 2002). Although CentO arrays are often discontinuous, owed presumably to the invasion/intermingling of retrotransposons, the arrangement of CentO in *Cen8* can be generalized as a single major array of approximately 60 kb within the CENH3 binding region, with a few smaller arrays residing outside the binding region. CRRs are the most abundant of the retrotransposon families found in *Cen8*, accounting for 20 of the 28 elements found in the CENH3 binding region (Nagaki et al. 2004). CRRs are distributed across a broader range of the centromere than the satellite repeats, and are found scattered throughout the centromere and even pericentromeric regions (Nagaki et al. 2005b).

Although studies of centromere evolution are still in their infancy, they have provided some accounts of an evolving centromere that include descriptions of neocentromere formation in humans (Lo et al. 2001a), characterizations of conserved centromere repeats in grasses (Miller et al. 1998; Presting et al. 1998; Zhong et al. 2002), and reports of centromere repositioning in primates (Cardone et al. 2006). Mature centromeres generally harbor long tracts of satellite DNA, are heavily invaded with retrotransposons, and are devoid of both genes and the ability to support transcription (Schueler et al. 2001). It is therefore surprising to learn that *Cen8* not only contains limited amounts of CentO satellite DNA, but perhaps more strikingly, active genes as well. Within the 750-kb CENH3 core domain of *Cen8*, a total of 16 genes were found with detectable transcript levels (Nagaki et al. 2004; Yan et al. 2005). The overall data suggest that rice *Cen8* may represent an intermediate or transitional stage in the development of a mature centromere.

Recent in-depth sequence analyses of both CentO and CRR contained within *cen8* have revealed opposing forces governing the dynamics of centromere development. Most notably, the recent and significant amount of segmental duplications of both satellite and CRRs found in *cen8* appears to be driving the expansion of the centromere (Ma and Bennetzen 2006; Ma and Jackson 2006). What may provide the most robust and accurate approach for understanding centromere dynamics and evolution is the comparative genomics of orthologous centromeres across related species. Preliminary analyses utilizing both the emerging BAC end sequence and physical mapping data between *cen8* of *O. sativa* and a wild progenitor (*O. brachyantha*) have already revealed significant rearrangements following their divergence less than 10 million years ago (Ge et al. 1999). *O. brachyantha* has completely lost any

remnant of the CentO satellite and has replaced it with another satellite, CentO-F, which is not found in any of the other 17 rice species examined (Lee et al. 2005; Ma et al. 2007). Furthermore, there is also discordance between the type and amount of retrotransposons harbored in the centromere of each of the two species, with *O. brachyantha* containing markedly fewer retrotransposons, some of which are again unique to the species (Lee et al. 2005; Ma et al. 2007). Finally, an inversion spanning one edge of the centromere in *O. brachyantha* provides evidence that the relative chromosomal position of this centromere has actually shifted with respect to its position in *O. sativa* (Ma et al. 2007). Thus, the changes seen within a span of less than 10 millions years suggest that centromeres are highly dynamic and are able to evolve quite rapidly at the sequence level.

6.4.3 Neocentromere

Neocentromeres are found on abnormal chromosomes that lack an original centromeric region containing a centromeric repeat, usually due to a deletion. To compensate for the absence, these anomalous chromosomes are shown to have generated a new centromere at a novel and non-centromeric region of the chromosome (Voullaire et al. 1993; Barry et al. 2000). Typically the neocentromere phenomenon is seen in patients with abnormal karyotypes, and studies within these groups have shown that centromeric proteins are localized at the neocentromere (Lo et al. 2001a, b; Alonso et al. 2007). These neocentromeres are frequently stable, and can faithfully be transmitted through several generations of meiotic divisions (Marshall et al. 2008).

In plants, the term "neocentromere" had been to describe subtelomeric heterochromatin regions in maize and rye that behave as centromere-like regions during meiosis rather than mitosis (Viinikka 1985; Yu et al. 1997). In maize, these heterochromatic regions are composed of non-centromeric repetitive DNA sequences, and lack centromeric proteins (Peacock et al. 1981; Dawe et al. 1999; Zhong et al. 2002). Recently, a neocentromere that is characteristically very similar to human neocentromeres was reported in barley (Nasuda et al. 2005). A line carrying the short arm of 7H chromosomes (7HS) was isolated from progenitors of a bread-wheat line carrying a pair of barley 7H chromosomes and a gametocidal chromosome 2C. Although the 7HS chromosome lost the barley centromeric repetitive DNA after a recombination with an unknown part (but non-centromeric) of the wheat chromosomes and was also subjected to some deletions, the reconstructed barley chromosome, nevertheless, still possess centromeric proteins and centromere functions. These situations are more similar to those involving human neocentromeres rather than the previously defined plant neocentromeres. However, the genesis of the barley neocentromere might differ from that of the human neocentromere. Although studies of human neocentromeres indicated that they are generated at new positions in a single step, the barley neocentromere appears to have shifted several times along the chromosomal arm region during the deletion steps to finally arrive at the current position.

6.4.4 Dicentric Chromosome

Inactivation of one of the two centromeres in dicentric chromosomes represents another epigenetic phenomenon seen in centromeric regions. Dicentric chromosomes possess two centromeric regions on a chromosome that can either be generated by a translocation between two chromosomes or a fusion between two sister chromatids. If both centromeres on a dicentric chromosome are active, the dicentric chromosome is usually unstable (McClintock 1939; Koshland et al. 1987; Lukaszewski 1995). Consequently, each centromere can be captured by spindle fibers eminating from different poles. As a result, the chromosome is pulled to different poles causing chromosome breakage. To avoid the breakage, one of the two centromeres on dicentric chromosomes is usually inactivated, leaving only one centromere to function as the active centromere (Sullivan and Schwartz 1995). Centromeric proteins disappear from the inactivated centromeres in this situation. Mechanisms responsible for the inactivation in dicentric chromosomes are not known for any organism. In addition, dicentric chromosomes with two active centromeres that are closely located are also stably transmittable (Koshland et al. 1987; Sullivan 1998). In this situation, these centromeres may be coordinated as a "single" centromere if the active centromeres are close enough. However, the acceptable distances between two active centromeres on dicentric chromosomes are quite different between species (1 kb to 12 Mb) (Koshland et al. 1987; Sullivan 1998).

Unstable dicentric chromosomes have been reported in maize (McClintock 1939); however, the first transmissible dicentric chromosome, among plants, was reported in wheat (Sears and Camara 1952). Interestingly, active state of one of the two centromeres on the wheat dicentric chromosome was epigenetically silenced. Specifically, the centromere was almost always active on unpaired univalents, but inactive on paired bivalents (Sears and Camara 1952). Recently, six dicentric chromosomes were characterized in maize (Han et al. 2006). Although the dicentric chromosomes showed two centromeric signals in FISH with maize centromeric DNA sequences, the dicentric chromosomes possessed only one CENH3 positive site as revealed by immunostaining with an anti-CENH3 antibody. Observations using a translocation line from B chromosome in maize revealed that the centromere from B chromosome in a dicentric chromosome was inactive for more than four generations (Han et al. 2006).

6.4.5 Holocentric Chromosome

A centromere is usually constructed at a primary constriction of chromosomes. However, a special type of chromosome is observed in some species in which the active centromere spans almost the entire length of the chromosome. This special type of chromosome is called a holocentric chromosome. These species (nematodes, insects, monocotyledons, and dicotyledons) are scattered among different biological kingdoms and yet most of their close relatives have the more conventional, localized

centromere (Kondo and Lavarack 1984; Pimpinelli and Goday 1989). Based on these facts, we can deduce that such holocentric species were derived independently from species possessing a localized centromere. Reorganization from a localized centromere chromosome to a holocentric chromosome represents a big change and involves a significant risk. It is interesting to note that such big changes have occurred several times and have become fixed in different phylogenetic lineages.

Although some holocentric species have been reported in different phylogenetic lines, *Caenorhabditis elegans* has been the most intensively studied at the molecular level (Dernburg 2001). In plants, holocentric species have been investigated only recently at the molecular level (Gernand et al. 2003; Haizel et al. 2005; Nagaki et al. 2005a). The centromere protein CENH3 (LnCENH3) was found in holocentric chromosomes of *Luzula nivea* (Nagaki et al. 2005a). Immunostaining with anti-LnCENH3 antibody revealed a linear-shaped centromere on mitotic metaphase chromosomes (Nagaki et al. 2005a). Although the amount of CENH3 is usually constant throughout the cell cycle in localized-centromere species, the amount varied during the cell cycles in *L. nivea*. The amount of LnCENH3 is limited to small areas during interphase and then increases from prometaphase to metaphase, resulting in the linear-shaped centromeres in metaphase. Subsequently, the amount of LnCENH3 decreases from telophase to interphase. Interestingly, similar changing patterns were observed for CENP-A and CENP-C in *C. elegans* (Buchwitz et al. 1999; Moore and Roth 2001; Oegema et al. 2001). These findings suggest that similar systems producing holocentric chromosomes were developed independently among different species at different points in evolutionary time. The increasing pattern may represent an inevitable way of combining gene expression at interphase and constructing holocentric chromosomes in metaphase. However, mechanisms controlling the loading of centromere proteins are not known for any holocentric species. Intriguingly, *C. elegans* lacks one of the two condensin subfamilies, and it has been suggested that the lack of a condensin subfamily is the cause of the change from a localized to holocentric centromere (Ono et al. 2003).

6.5 Centromere Modification

6.5.1 General Remarks

In general, centromeric and paracentric regions are epigenetically modified by histone acetylations, methylations, and phosphorylations at different amino acid positions, and by DNA methylations. For instance, dimethylations of histone H3 at lysine 4 (H3K4me2) and acetylations of histone H4 (H4Ac) are characteristic of euchromatin, while dimethylations of histone H3 at lysine 9 (H3K9me2) and DNA methylations at cytosine bases are associated with heterochromatin (Turner 2002). In human and *D. melanogaster* centromeres, centromeric chromatins including CENH3 were also associated with H3K4me2, while pericentromeric regions located outside of the centromeric chromatins possessed H3K9me2 (Sullivan and Karpen 2004). A similar co-localization pattern of CENH3 chromatin and these dimethylated histone H3s was

also observed in *S. pombe* (Cam et al. 2005). In addition, phosphorylations of histone H3 at Ser10 and Ser28 have been observed in mitotic cells among animals and plants. In mammals, the phosphorylation of histone H3 at Ser10 begins in pericentric heterochromatins and spreads to the chromosome arms during metaphase (Hendzel et al. 1997). However, in plants, the phosphorylation of histone H3 at Ser10 and Ser28 is reduced in the pericentric heterochromatins from prophase stage to telophase stage in mitosis, but spreads along the entire chromosome in meiosis (Houben et al. 1999; Kaszas and Cande 2000; Gernand et al. 2003). CENH3 is also phosphorylated in its N-terminal tail (Zeitlin et al. 2001). In humans, Ser7 of CENP-A, the Ser10 counterpart of the canonical histone H3, is phosphorylated in a pattern similar to the phosphorylation of histone H3 (Zeitlin et al. 2001). In maize, Ser50 of CENH3, Ser28 counterpart of the canonical histone H3, is phosphorylated in a pattern similar to that of histone H3 (Zhang et al. 2005). DNA methylations also works as an epigenetic marker for heterochromatins. Since the heavy methylation status of centromeric repetitive DNA sequences has been reported for various organisms (Martinez-Zapater et al. 1986; Miniou et al. 1997; Dong et al. 1998), it was believed that centromeres have highly methylated regions throughout their entire length. However, it is unclear whether all or just portions of centromere sequences, within a defined region, are methylated.

6.5.2 Arabidopsis

Immunostaining using labeled anti-HTR12 and anti-methylated cytosine antibodies on extended chromatin fiber uncovered the hypomethylation status of *Arabidopsis* centromeres (Zhang et al. 2008). CENH3 binding regions of *Arabidopsis* centromeres possess hypomethylated 180-bp repeat family sequences, while the DNA located outside the binding region was highly methylated. Additionally, immunostaining using anti-HTR12 and anti-Lys9 dimethylated histone H3 antibodies showed patterns similar to those obtained with anti-HTR12 and anti-methylated cytosine antibodies. ChIP cloning using anti-HTR12 antibodies uncovered two distinct types of 180-bp repeat family sequences (Zhang et al. 2008). One is centromeric and is among the precipitated sequences, while the other is pericentromeric in location and not found in the precipitated sequences. Distribution patterns of CG and CNG methylation sites differed between these two types of sequences. These results have revealed distinct epigenetic features of the CENH3-associated "centromeric chromatin" as compared to the pericentromeric heterochromatin.

6.5.3 Rice

The fully sequenced centromeres of chromosome 8 and 4 in rice provide an exceptional source for an analysis of the structural features of active chromosomes, both epigenetically and at the sequence level (Nagaki et al. 2004; Wu et al. 2004; Zhang et al. 2004). Although previously reported human, *D. melanogaster*, and *S. pombe* centromeres

have CENH3 and H3K4me2, rice *Cen8* appears to be mostly enriched for H3K9me2 (Nagaki et al. 2004). Both centromere 8 and centromere 3 contain genes with detectable transcripts, and the precise manual annotations of these centromere regions have revealed that local modifications marking euchromatin such as H3K4me2 and H4Ac enrich the genes within the centromere (Yan et al. 2005, 2006). The patterns of the modified histones associated with the centromeric genes appear to be similar to those seen in the pericentromeric regions. Further genome-wide analysis may reveal whether rice centromeres are associated with an unique combination of histone modification patterns reported in animal centromeres.

DNA methylation studies of centromeres, while still preliminary, have provided some evidence that epigenetic DNA modifications are found in the centromeres of rice. The level of methylation across the centromeres of rice is generally discontinuous, with the highest enrichment distributed broadly across the CentO satellite repeats (Li et al. 2008). Intriguingly, CRRs showed little evidence of being methylated (Li et al. 2008). While this finding coincides with the fact that some CR elements are actively transcribed in the centromere (Neumann et al. 2007), the hypomethylation of CRRs may prove to be a unique feature for this family of retrotransposons, given that most retrotransposon families located outside the centromere are typically hypermethylated.

6.6 Minichromosomes and Artificial Chromosomes

6.6.1 Minichromosomes

Minichromosomes are considerably smaller than their native equivalent and also harbor a much more succinct centromere. These characteristics make them a highly useful tool for determining the functional domains of centromeres. Initial studies focused on budding yeast *S. cerevisiae* that has very short (~125 bp) centromeres (Cottarel et al. 1989). However, utilization of the minichromosome approach is more informative in the eukaryotes, which have longer and more complex centromeres, such as *S. pombe* (Hahnenberger et al. 1989), *D. melanogaster* (Sun et al. 2003), and humans (Auriche et al. 2001). These minichromosomes have been produced by homologous recombination, γ irradiation, or telomere-insertion-induced truncation.

Minichromosomes have been found also in higher plants: the PRO1 chromosome of *Beta procumbens* (Gindullis et al. 2001a, b), the B chromosome of *Brachycome dichromosomatica* (Houben et al. 1997), and the midget chromosome of *Secale cereale* (Murata et al. 1992). In maize, various types of minichromosomes have been created by the breakage-fusion-bridge (BFB) cycle from a translocation between the B chromosome and chromosome 9 short-arm in maize (Zheng et al. 1999; Kato et al. 2005). The overall size and structure of the mini-B-chormosomes were different from one another, although the exact sizes of the minichromosomes or their centromere sizes were not determined.

In *A. thaliana*, the minichromosomes found were considerably smaller than those in maize and other plant species: an expected result, given the size of the native chromosomes are 25.3–38.0 Mb (Hosouchi et al. 2002, Murata et al. 2006, 2008). An *Arabidopsis* minichromosome, named mini-4S, is thought to originate from a breakage of the centromere of chromosome 4, and thus contains the whole short arm and the truncated centromere of chromosome 4 (Murata et al. 2006). The size of this mini-4S chromosome and the amount of the major centromeric satellite (180-bp family) are estimated to be ~7.5 and ~1 Mb, respectively. Although these sizes are approximately one-third of the native chromosome 4, the transmission rates of mini-4S to the next generation were almost coincident with the rate of chromosome 4 in the primary trisomic condition (Koornneef and Van der Veen 1983). The transmission rate of mini-4S is considerably higher in the Columbia-ecotype background than in the Landsberg-*erecta*-ecotype. This suggests that the minichromosome stability is controlled genetically, as shown in *S. pombe*. Recently, two other minichromosomes (mini-α and mini-δ) were found in a transgenic *Arabidopsis* plant (Murata et al. 2008). Both minichromosomes originated from a short arm of chromosome 2. The size of mini-α is ~8 Mb, almost the same as that of mini-4S, but the amount of the centromeric major satellite (180-bp family) is ~0.7 Mb, which is less than that of mini-4S. Mini-δ is a ring chromosome carrying two copies of the 180-bp repeat cluster and of the pericentric region, the estimated size of which is 3.7 Mb in total. Although it lacks the terminal region including telomere repeats, it is stably transmitted to the next generations but only through the male gametes. Each of the two 180-bp clusters (ca. 500 kb in length) is thought to allow normal centromere function because the centromere-specific histone H3 (HTR12) was detected on both clusters. These results indicate that 500 kb of the 180-bp array is sufficient to form a functional kinetochore.

A lower size limit (5% of genome size) is suggested as the minimum amount to promote stable transmission through meiosis (Schubert 2001). The size (7.5–8 Mb) of mini-4S and mini-α is consistent with this hypothesis, whereas the size of mini-δ is inconsistent, as is PRO1 minichromosome, the size of which is only 6–9 Mb. The size of centromeres and/or shape of chromosome (linear vs. circular) may also be important for the minichromosome stability.

6.6.2 Artificial Chromosomes

In addition to the centromere, an eukaryotic chromosome contains at least two more functional elements: telomere and origin of replication (ori). These three elements are required for chromosome maintenance and stability during cell division. This requirement was first shown in budding yeast by constructing artificial chromosomes (YAC) from the DNA molecules assembled in vitro (Murray and Szostak 1983). This type of approach for the construction of artificial chromosomes, called "the bottom-up approach," is also effective in constructing mammalian or human artificial chromosomes (MAC or HAC) (Harrington et al. 1997; Ikeno et al. 1998).

More recently in maize, a circular centromeric DNA constructed in vitro was demonstrated to be heritable when introduced into the embryonic tissues (Carlson et al. 2007), but the evidence supporting the de novo formation of artificial chromosomes have yet to be reported (Houben et al. 2008). Presently, it is difficult to introduce large centromere-specific repetitive DNA molecules into plant cells efficiently. Therefore, the minichromosome approach described earlier is an alternative choice for constructing artificial chromosomes in plants.

Chromosomal breakage caused by the introduction of cloned telomere DNA was first applied in human cells (Itzhaki et al. 1992), and since then it has been used to generate the minichromosomes (Heller et al. 1996; Saffery et al. 2001). Recently, telomere-mediated chromosomal truncation was shown to be feasible approach in maize for generating A and B minichromosomes (Yu et al. 2007). Since these minichromosomes are manipulated to contain a *Lox* P site, site-specific recombination of desired genes is possible with Cre recombinase. Very recently, it was reported that minichromosomes were formed in maize by introducing the DNA molecule containing native centromere segment, ori, and telomere repeats (Ananiev et al. 2008).

6.7 Concluding Remarks

Investigations of plant centromeric components and centromere structures during this decade have broadened our knowledge of plant centromeres. However, most of the parts of plant centromeres are still unknown. More direct interrogation of plant centromere complexes, mutants of plant centromere proteins, and artificial chromosomes will lead to a greater understanding of plant centromeres.

References

Alfenito MR, Birchler JA (1993) Molecular characterization of a maize B chromosome centric sequence. Genetics 135:589–597

Alonso A, Fritz B, Hasson D, Abrusan G, Cheung F, Yoda K, Radlwimmer B, Ladurner AG, Warburton PE (2007) Co-localization of CENP-C and CENP-H to discontinuous domains of CENP-A chromatin at human neocentromeres. Genome Biol 8:R148

Amor DJ, Kalitsis P, Sumer H, Choo KH (2004) Building the centromere: from foundation proteins to 3D organization. Trends Cell Biol 14:359–368

Ananiev EV, Phillips RL, Rines HW (1998) Chromosome-specific molecular organization of maize. (*Zea mays* L.) centromeric regions. Proc Natl Acad Sci USA 95:13073–13078

Ananiev EV, Svitashev S, Wu C, Chamberlin MA, Gordon-Kamm W, Schwartz C, Sturdevant M, Tingey S (2008) Artificial chromosome construction in *Zea mays* L. In: XX Internat. Congress Genetics, July 11–17, 2008, Berlin, Germany, p. 346

Aragon-Alcaide L, Miller T, Schwarzacher T, Reader S, Moore G (1996) A cereal centromeric sequence. Chromosoma 105:261–268

Auriche C, Donini P, Ascenzioni F (2001) Molecular and cytological analysis of a 5.5 Mb minichromosome. EMBO Rep 2:102–107

Barry A, Bateman M, Howman E, Cancilla M, Tainton K, Irvine D, Saffery R, Choo K (2000) The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification. Genome Res 10:832–838

Black BE, Bassett EA (2008) The histone variant CENP-A and centromere specification. Curr Opin Cell Biol 20:91–100

Blower MD, Sullivan BA, Karpen GH (2002) Conserved organization of centromeric chromatin in flies and humans. Dev Cell 2:319–330

Buchwitz BJ, Ahmad K, Moore LL, Roth MB, Henikoff S (1999) A histone-H3-like protein in *C. elegans*. Nature 401:547–548

Cam HP, Sugiyama T, Chen ES, Chen X, FitzGerald PC, Grewal SI (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. Nat Genet 37:809–819

Cardone MF, Alonso A, Pazienza M, Ventura M, Montemurro G, Carbone L, de Jong PJ, Stanyon R, D'Addabbo P, Archidiacono N, She X, Eichler EE, Warburton PE, Rocchi M (2006) Independent centromere formation in a capricious, gene-free domain of chromosome 13q21 in Old World monkeys and pigs. Genome Biol 7:R91

Carlson SR, Rudgers GW, Zieler H, Mach JM, Luo S, Grunden E, Krol C, Copenhaver GP, Preuss D (2007) Meiotic transmission of an *in vitro*-assembled autonomous maize minichromosome. PLoS Genet 3:1965–1974

Chang SB, Yang TJ, Datema E, van Vugt J, Vosman B, Kuipers A, Meznikova M, Szinay D, Lankhorst RK, Jacobsen E, de Jong H (2008) FISH mapping and molecular organization of the major repetitive sequences of tomato. Chromosome Res 16(7):919–933

Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell 14:1691–1704

Cheng Z-J, Murata M (2003) A centromeric tandem repeat family originating from a part of Ty3/gypsy-retroelement in wheat and its relatives. Genetics 164:665–672

Choo KH (2001) Domain organization at the centromere and neocentromere. Dev Cell 1:165–177

Choo KHA (1997) The centromere. Oxford University Press, Oxford

Cooper JL, Henikoff S (2004) Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol 21:1717–1718

Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD, McCombie WR, Martienssen RA, Marra M, Preuss D (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science 286:2468–2474

Cottarel G, Shero JH, Hieter P, Hegemann JH (1989) A 125-base-pair CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae*. Mol Cell Biol 9:3342–3349

Dawe RK, Henikoff S (2006) Centromeres put epigenetics in the driver's seat. Trend in Biochem Sci 31:662–669

Dawe RK, Reed LM, Yu HG, Muszynski MG, Hiatt EN (1999) A maize homolog of mammalian CENPC is a constitutive component of the inner kinetochore. Plant Cell 11:1227–1238

Dernburg AF (2001) Here, there, and everywhere: kinetochore function on holocentric chromosomes. J Cell Biol 153:F33–F38

Dong F, Miller JT, Jackson SA, Wang GL, Ronald PC, Jiang J (1998) Rice. (*Oryza sativa*) centromeric regions consist of complex DNA. Proc Natl Acad Sci USA 95:8135–8140

Du Y, Dawe RK (2007) Maize NDC80 is a constitutive feature of the central kinetochore. Chromosome Res 15:767

Earnshaw WC, Rothfield N (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma 91:313–321

Entani T, Iwano M, Shiba H, Takayama S, Fukui K, Isogai A (1999) Centromeric localization of an S-RNase gene in *Petunia hybrida* Vilm. Theor Appl Genet 99:391–397

Francki MG (2001) Identification of Bilby, a diverged centromeric Ty1-copia retrotransposon family from cereal rye. (*Secale cereale* L.). Genome 44:266–274

Fukagawa T, Brown WRA (1997) Efficient conditional mutation of the vertebrate CENP-C gene. Hum Mol Genet 6:2301–2308

Fukui KN, Suzuki G, Lagudah ES, Rahman S, Appels R, Yamamoto M, Mukai Y (2001) Physical arrangement of retrotransposon-related repeats in centromeric regions of wheat. Plant Cell Physiol 42:189–196

Furuyama S, Biggins S (2007) Centromere identity is specified by a single centromeric nucleosome in budding yeast. Proc Natl Acad Sci USA 104:14706–14711

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Ge S, Sang T, Lu BR, Hong DY (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc Natl Acad Sci USA 96:14400–14405

Gernand D, Demidov D, Houben A (2003) The temporal and spatial pattern of histone H3 phosphorylation at serine 28 and serine 10 is similar in plants but differs between mono- and polycentric chromosomes. Cytogenet Genome Res 101:172–176

Gindullis F, Dechyeva D, Schmidt T (2001a) Construction and characterization of a BAC library for the molecular dissection of a single wild beet centromere and sugar beet. (*Beta vulgaris*) genome analysis. Genome 44:846–855

Gindullis F, Desel C, Galasso I, Schmidt T (2001b) The large-scale organization of the centromeric region in *Beta* species. Genome Res 11:253–265

Goel S, Raina SN, Ogihara Y (2002) Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of nuclear ribosomal DNA in the Phaseolus-Vigna complex. Mol Phylogenet Evol 22:1–19

Goshima G, Saitoh S, Yanagida M (1999) Proper metaphase spindle length is determined by centromere proteins Mis12 and Mis6 required for faithful chromosome segregation. Genes Dev 13:1664–1677

Goshima G, Kiyomitsu T, Yoda K, Yanagida M (2003) Human centromere chromatin protein hMis12, essential for equal segregation, is independent of CENP-A loading pathway. J Cell Biol 160:25–39

Hahnenberger KM, Baum MP, Polizzi CM, Carbon J, Clarke L (1989) Construction of functional artificial minichromosomes in the fission yeast *Schizosaccharomyces pombe*. Proc Natl Acad Sci USA 86:577–581

Haizel T, Lim YK, Leitch AR, Moore G (2005) Molecular analysis of holocentric centromeres of *Luzula* species. Cytogenet Genome Res 109:134–143

Han F, Lamb JC, Bircheler JA (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. Proc Natl Acad Sci USA 103:3238–3243

Harrington JJ, Bokkelen GV, Mays RW, Gustashaw K, Willard HF (1997) Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. Nat Genet 15:345–355

Harrison GE, Heslop-Harrison JS (1995) Centromeric repetitive DNA sequences in the genus *Brassica*. Theor Appl Genet 90:157–165

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148:479–494

Heller K, Kilian A, Piatyszek MA, Kleinhofs A (1996) Telomerase activity in plant extracts. Mol Gen Genet 252:342–345

Hendzel MJ, Wei Y, Mancini MA, Van Hooser A, Ranalli T, Brinkley BR, Bazett-Jones DP, Allis CD (1997) Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. Chromosoma 106:348–360

Henikoff S, Ahmad K, Platero JS, van Steensel B (2000) Heterochromatic deposition of centromeric histone H3-like proteins. Proc Natl Acad Sci USA 97:716–721

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098–1102

Heslop-Harrison JS, Brandes A, Schwarzacher T (2003) Tandemly repeated DNA sequences and centromeric chromosomal regions of *Arabidopsis* species. Chromosome Res 11:241–253

Hizume M, Shibata F, Maruyama Y, Kondo T (2001) Cloning of DNA sequences localized on proximal fluorescent chromosome bands by microdissection in *Pinus densiflora* Sieb. & Zucc. Chromosoma 110:345–351

Hosouchi T, Kumekawa N, Tsuruoka H, Kotani H (2002) Physical map-based sizes of the centromeric regions of *Arabidopsis thaliana* chromosomes 1, 2, and 3. DNA Res 9:117–121

Houben A, Schubert I (2003) DNA and proteins of plant centromeres. Curr Opin Plant Biol 6:554–560

Houben A, Guttenbach M, Kress W, Pich U, Schubert I, Schmid M (1995) Immunostaining and interphase arrangement of field bean kinetochores. Chromosome Res 3:27–31

Houben A, Leach CR, Verlin D, Rofe R, Timmis JN (1997) A repetitive DNA sequence common to the different B chromosomes of the genus *Brachycome*. Chromosoma 106:513–519

Houben A, Wako T, Furushima-Shimogawara R, Presting G, Kunzel G, Schubert I, Fukui K (1999) The cell cycle dependent phosphorylation of histone H3 is correlated with the condensation of plant mitotic chromosomes. Plant J 18:675–679

Houben A, Schroeder-Reiter E, Nagaki K, Nasuda S, Wanner G, Murata M, Endo TR (2007) CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley. Chromosoma 116:275–283

Houben A, Dawe RK, Jiang J, Schubert I (2008) Engineered plant minichromosomes: A bottom-up success. Plant Cell 20:8–10

Hudakova S, Michalek W, Presting GG, ten Hoopen R, dos Santos K, Jasencakova Z, Schubert I (2001) Sequence organization of barley centromeres. Nucleic Acids Res 29:5029–5035

Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H (1998) Construction of YAC-based mammalian artificial chromosomes. Nat Biotechnol 16:431–439

Itzhaki JE, Barnett MA, MacCarthy AB, Buckle VJ, Brown WR, Porter AC (1992) Targeted breakage of a human chromosome mediated by cloned human telomeric DNA. Nat Genet 2:238–287

Jiang J, Nasuda S, Dong F, Scherrer CW, Woo S, Wing RA, Gill BS, Ward DC (1996) A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proc Natl Acad Sci USA 93:14210–14213

Jin W, Melo JR, Nagaki K, Talbert PB, Henikoff S, Dawe RK, Jiang J (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. Plant Cell 16:571-581

Kalitsis P, Fowler KJ, Earle E, Hill J, Choo KH (1998) Targeted disruption of mouse centromere protein C gene leads to mitotic disarray and early embryo death. Proc Natl Acad Sci USA 95:1136–1141

Kamm A, Schmidt T, Heslop-Harrison JS (1994) Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum*. Mol Gen Genet 244:420–425

Kamm A, Galasso I, Schmidt T, Heslop-Harrison JS (1995) Analysis of a repetitive DNA family from *Arabidopsis arensa* and relationship between *Arabidopsis* species. Plant Mol Biol 27:853–862

Kaszas E, Cande WZ (2000) Phosphorylation of histone H3 is correlated with changes in the maintenance of sister chromatid cohesion during meiosis in maize, rather than the condensation of the chromatin. J Cell Sci 113:3217–3226

Kato A, Zheng YZ, Auger DL, Phelps-Durr T, Bauer MJ, Lamb JC, Birchler JA (2005) Minichromosomes derived from the B chromosome of maize. Cytogenet Genome Res 109:156–165

Kawabe A, Nasuda S (2005) Structure and genomic organization of centromeric repeats in *Arabidopsis* species. Mol Gen Genomics 272:593–602

Kawabe A, Nasuda S, Charlesworth D (2006) Duplication of centromeric histone H3. (HTR12) gene in *Arabidopsis halleri* and *A. lyrata*, plant species with multiple centromeric satellite sequences. Genetics 174:2021–2032

Kikuchi S, Kishii M, Shimizu M, Tsujimoto H (2005) Centromere-specific repetitive sequences from *Torenia*, a model plant for interspecific fertilization, and whole-mount FISH of its interspecific hybrid embryos. Cytogenet Genome Res 109:228–235

Kishii M, Nagaki K, Tsujimoto H (2001) A tandem repetitive sequence located in the centromeric region of common wheat. (*Triticum aestivum*) chromosomes. Chromosome Res 9:417–428

Kondo K, Lavarack PS (1984) A cytotaxonomic study of some Australian species of *Drosera* L. (Droseraceae). Botanical J Linnean Soc 88:317–333

Koornneef M, Van der Veen JH (1983) The trisomics of *Arabidopsis thaliana*. (L.) Heynh. and the location of linkage groups. Genetica 61:41–46

Koshland D, Rutledge L, Fitzgerald-Hayes M, Hartwell LH (1987) A genetic analysis of dicentric minichromosomes in *Saccharomyces cerevisiae*. Cell 48:801–812

Kumekawa N, Hosouchi T, Tsuruoka H, Kotani H (2000) The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 5. DNA Res 7:315–321

Kumekawa N, Hosouchi T, Tsuruoka H, Kotani H (2001) The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 4. DNA Res 8:285–290

Leach CR, Donald TM, Franks TK, Spiniello SS, Hanrahan CF, Timmis JN (1995) Organisation and origin of a B chromosome centromeric sequence from *Brachycome dichromosomatica*. Chromosoma 103:708–714

Lee H, Zhang W, Langdon T, Jin W, Yan H, Cheng Z, Jiang J (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. Proc Natl Acad Sci USA 102:11793–11798

Li X, Wang X, He K, Ma Y, Su N, He H, Stolc V, Tongprasit W, Jin W, Jiang J, Terzaghi W, Li S, Deng XW (2008) High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. Plant Cell 20:259–276

Liu Z, Yue W, Li D, Wang RR, Kong X, Lu K, Wang G, Dong Y, Jin W, Zhang X (2008) Structure and dynamics of retrotransposons at wheat centromeres and pericentromeres. Chromosoma 117:445–456

Lo AW, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E, Magliano DJ, Choo KH (2001a) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. EMBO J 20:2087–2096

Lo AW, Magliano DJ, Sibson MC, Kalitsis P, Craig JM, Choo KH (2001b) A novel chromatin immunoprecipitation and array. (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. Genome Res 11:448–457

Lukaszewski AJ (1995) Chromatid and chromosome type breakage-fusion-bridge cycles in wheat. (*Triticum aestiuum* L.). Genetics 140:1069–1085

Ma J, Bennetzen JL (2006) Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice. Proc Natl Acad Sci USA 103:383–388

Ma J, Jackson SA (2006) Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. Genome Res 16:251–259

Ma J, Wing RA, Bennetzen JL, Jackson SA (2007) Evolutionary history and positional shift of a rice centromere. Genetics 177:1217–1220

Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. Genetics 157:1293–1298

Maluszynsak J, Heslop-Harrison JS (1991) Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. Plant J 1:159–166

Marshall OJ, Chueh AC, Wong LH, Choo KH (2008) Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am J Hum Genet 82:261–282

Martinez-Zapater J, Estelle MA, Somerville CR (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. Mol Gen Genet 204:417–423

McClintock B (1939) The behavior in successive muclear divisions of a chromosome broken at meiosis. Genetics 25:405–416

Meluh PB, Yang P, Glowczewski L, Koshland D, Smith MM (1998) Cse4p is a component of the core centromere of Saccharomyces cerevisiae. Cell 94:607–613

Miller JT, Jackson SA, Nasuda S, Gill BS, Wing RA (1998) Cloning and characterization of a centromere-specific repetitive DNA element from *Sorghum bicolor*. Theor Appl Genet 96:832–839

Miniou P, Jeanpierre M, Bourc'his D, Coutinho Barbosa AC, Blanquet V, Viegas-Peuignot E (1997) Alpha-satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. Hum Genet 99:738–745

Mole-Bajer J, Bajer AS, Zinkowski RP, Balczon RD, Brinkley BR (1990) Autoantibodies from a patient with scleroderma CREST recognized kinetochores of the higher plant *Haemanthus*. Proc Natl Acad Sci USA 87:3599–3603

Moore LL, Roth MB (2001) HCP-4, a CENP-C-like protein in *Caenorhabditis elegans*, is required for resolution of sister centromeres. J Cell Biol 153:1199–1208

Murata M, Nakata N, Yasumuro Y (1992) Origin and molecular structure of a midget chromosome in a common wheat carrying rye cytoplasm. Chromosoma 102:27–31

Murata M, Ogura Y, Motoyoshi F (1994) Centromeric repetitive sequences in *Arabidopsis thaliana*. Jpn J Genet 69:361–370

Murata M, Shibata F, Yokota E (2006) The origin, meiotic behavior, and transmission of a novel minichromosome in *Arabidopsis thaliana*. Chromosoma 115:311–319

Murata M, Yokota E, Shibata F, Kashihara K (2008) Functional analysis of the *Arabidopsis* centromere by T-DNA insertion-induced centromere breakage. Proc Natl Acad Sci USA 105:7511–7516

Murray AW, Szostak JW (1983) Construction of artificial chromosomes in yeast. Nature 305:189–193

Nagaki K, Murata M (2005) Characterization of CENH3 and centromere-associated DNA sequences in sugarcane. Chromosome Research 13:195–203

Nagaki K, Tsujimoto H, Sasakuma T (1998) A novel repetitive sequence of sugar cane, SCEN family, locating on centromeric regions. Chromosome Res 6:295–302

Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang J (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics 163:1221–1225

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145

Nagaki K, Kashihara K, Murata M (2005a) Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant *Luzula nivea*. Plant Cell 17:886–1893

Nagaki K, Neumann P, Zhang D, Ouyang S, Buell CR, Cheng Z, Jiang J (2005b) Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. Mol Biol Evol 22:845–855

Nasuda S, Hudakova S, Schubert I, Houben A, Endo T (2005) Stable barley chromosomes without centromeric repeats. Proc Natl Acad Sci USA 102:9842–9847

Neumann P, Yan H, Jiang J (2007) The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. Genetics 176:749–761

Nonomura KI, Kurata N (1999) Organization of the 1.9-kb repeat unit RCE1 in the centromeric region of rice chromosomes. Mol Gen Genet 261:1–10

Obuse C, Yang H, Nozaki N, Goto S, Okazaki T, Yoda K (2004) Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1. (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. Genes Cells 9:105–120

Oegema K, Desai A, Rybina S, Kirkham M, Hyman AA (2001) Functional analysis of kinetochore assembly in *Caenorhabditis elegans*. J Cell Biol 153:1209–1226

Ogura Y, Shibata F, Sato H, Murata M (2004) Characterization of a CENP-C homolog in *Arabidopsis thaliana*. Genes Genet Syst 79:139–144

Okada M, Cheeseman I, Hori T, Okawa K, McLeod I, Yates Jr, Desai A, Fukagawa T (2006) The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. Nat Cell Biol 8:446–457

Ono T, Losada A, Hirano M, Myers MP, Neuwald AF, Hirano T (2003) Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. Cell 115:109–121

Page BT, Wanous MK, Birchler JA (2001) Characterization of a maize chromosome 4 centromeric sequence. Evidence for an evolutionary relationship with the B chromosome centromere. Genetics 159:291–302

Palmer DK, O'Day K, Wener MH, Andrews BS, Margolis RL (1987) A 17-kD centromere protein. (CENP-A) copurifies with nucleosome core particles and with histones. J Cell Biol 104:805–815

Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci USA 88:3734–3738

Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ (1981) Highly repeated DNA sequence limited to knob heterochromatin in maize. Proc Natl Acad Sci USA 78:4490–4494

Pelissier T, Tutois S, Deragon JM, Tourmente S, Genestier S, Picard G (1995) *Athila*, a new retroelement from *Arabidopsis thaliana*. Plant Mol Biol 29:441–452
Pimpinelli S, Goday C (1989) Unusual kinetochores and chromatin diminution in Parascaris. Trends Genet 5:310–315
Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A Ty3/gypsy retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J 16:721–728
Preuss D, Rhee SY, Davis RW (1994) Tetrad analysis possible in *Arabidopsis* with mutation of the QUARTET. (QRT) genes. Science 264:1458–1460
Saffery R, Wong LH, Irvine DV, Bateman MA, Griffiths B, Cutts SM, Cancilla MR, Cendron AC, Stafford AJ, Choo KH (2001) Construction of neocentromere-based human minichromosomes by telomere-associated chromosomal truncation. Proc Natl Acad Sci USA 98:5705–5710
Saitoh H, Tomkiel J, Cooke CA, Ratrie Hr, Maurer M, Rothfield NF, Earnshaw WC (1992) CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. Cell 70:115–125
Sato H, Shibata F, Murata M (2005) Characterization of a Mis12 homologue in *Arabidopsis thaliana*. Chromosome Res 13:827–834
Saunders VA, Houben A (2001) The pericentromeric heterochromatin of the grass *Zingeria biebersteiniana*. (2n = 4) is composed of Zbcen1-type tandem repeats that are intermingled with accumulated dispersedly organized sequences. Genome 44:955–961
Schmidt T, Heslop-Harrison JS (1996) High-resolution mapping of repetitive DNA by *in situ* hybridization: molecular and chromosomal features of prominent dispersed and discretely localized DNA families from wild beet species *Beta procumbens*. Plant Mol Biol 30:1099–1114
Schmidt T, Metzlaff M (1991) Cloning and characterization of a *Beta vulgaris* satellite DNA family. Gene 101: 247–250
Schubert I (2001) Alteration of chromosome numbers by generation of minichromosomes – is there a lower limit of chromosome size for stable segregation. Cytogenet Cell Genet 93:175–181
Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. Science 294:109–115
Sears ER, Camara A (1952) A transmissible dicentric chromosome. Genetics 37:125–135
Shibata F, Murata M (2004) Differential localization of the centromere-specific proteins in the major centromeric satellite of *Arabidopsis thaliana*. J Cell Sci 117:2963–2970
Starr DA, Williams BC, Li Z, Etemad-Moghadam B, Dawe RK, Goldberg ML (1997) Conservation of the centromere/kinetochore protein ZW10. J Cell Biol 138:1289–1301
Sullivan B, Schwartz S (1995) Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. Hum Mol Genet 4:2189–2197
Sullivan BA (1998) Stable dicentric X chromosomes with two functional centromeres. Nat Genet 20:227–228
Sullivan BA, Karpen GH (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat Struct Mol Biol 11:1076–1083
Sun X, Le HD, Wahlstrom JM, Karpen GH (2003) Sequence analysis of a functional *Drosophila* centromere. Genome Res 13:182–194
Suzuki T, Ide N, Tanaka I (1997) Immunocytochemical visualization of the centromeres during male and female meiosis in *Lilium longiflorum*. Chromosoma 106:435–445
Takahashi K, Yamada H, Yanagida M (1994) Fission yeast minichromosome loss mutants mis cause lethal aneuploidy and replication abnormality. Mol Biol Cell 5:1145–1158
Talbert PB, Bryson TD, Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. J Biol 3:18
Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. Plant Cell 14:1053–1066
Tek AL, Jiang J (2004) The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. Chromosoma 113:77–83
ten Hoopen R, Manteuffel R, Dolezel J, Malysheva L, Schubert I (2000) Evolutionary conservation of kinetochore protein sequences in plants. Chromosoma 109:482–489

ten Hoopen R, Schleker T, Manteuffel R, Schubert I (2002) Transient CENP-E-like kinetochore proteins in plants. Chromosome Res 10:561–570

The Arabidopsis genome initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Turner BM (2002) Cellular memory and the histone code. Cell 111:285–291

Vermaak D, Hayden HS, Henikoff S (2002) Centromere targeting element within the histone fold domain of Cid. Mol Cell Biol 22:7553–7561

Viinikka Y (1985) Identification of the chromosome showing neocentric activity in rye. Theor Appl Genet 70:66–71

Voullaire LE, Slater HR, Petrovic V, Choo KHA (1993) A functional marker centromere with no detectable alpha-satellite, satellite-III, or CENP-B protein-activation of a latent centromere. American J Human Genet 52:1153–1163

Wevrick R, Willard HF (1989) Long-range organization of tandem arrays of alpha satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. Proc Natl Acad Sci USA 86:9394–9398

Wu J, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, Shibata M, Ito Y, Nakamura M, Sakaguchi M, Yoshihara R, Kobayashi H, Ito K, Karasawa W, Yamamoto M, Saji S, Katagiri S, Kanamori H, Namiki N, Katayose Y, Matsumoto T, Sasaki T (2004) Composition and structure of the centromeric region of rice chromosome 8. Plant Cell 16:967–976

Yan H, Ito H, Nobuta K, Ouyang S, Jin W, Tian S, Lu C, Venu RC, Wang GL, Green PJ, Wing RA, Buell CR, Meyers BC, Jiang J (2006) Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. Plant Cell 18:2123–2133

Yan H, Jin W, Nagaki K, Tian S, Ouyang S, Buell CR, Talbert PB, Henikoff S, Jiang J (2005) Transcription and histone modifications in the recombination-free region spanning a rice centromere. Plant Cell 17:3227–3238

Yang CH, Tomkiel J, Saitoh H, Johnson DH, Earnshaw WC (1996) Identification of overlapping DNA-binding and centromere-targeting domains in the human kinetochore protein CENP-C. Mol Cell Biol 16:3576–3586

Yu HG, Hiatt EN, Chan A, Sweeney M, Dawe RK (1997) Neocentromere-mediated chromosome movement in maize. J Cell Biol 139:831–840

Yu HG, Muszynski MG, Dawe RK (1999) The maize homologue of the cell cycle checkpoint protein MAD2 reveals kinetochore substructure and contrasting mitotic and meiotic localization patterns. J Cell Biol 145:425–435

Yu W, Han F, Gao Z, Vega JM, Birchler JA (2007) Construction and behavior of engineered minichromosomes in maize. Proc Natl Acad Sci USA 104:8924–8929

Zeitlin SG, Shelby RD, Sullivan KF (2001) CENP-A is phosphorylated by Aurora B kinase and plays an unexpected role in completion of cytokinesis. J Cell Biol 155:1147–1157

Zhang W, Lee HR, Koo DH, Jiang J (2008) Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. Plant Cell 20:25–34

Zhang X, Li X, Marshall JB, Zhong CX, Dawe RK (2005) Phosphoserines on maize centromeric histone H3 and histone H3 demarcate the centromere and pericentromere during chromosome segregation. Plant Cell 17:572–583

Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y, Wing RA, Han B (2004) Structural features of the rice chromosome 4 centromere. Nucl Acids Res 32:2023–2030

Zheng YZ, Roseman RR, Carlson WR (1999) Time course study of the chromosome-type breakage-fusion-bridge cycle in maize. Genetics 153:1435–1444

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

Index