



HAL
open science

Formation of the CenH3-independent holocentromere in Lepidoptera avoids active chromatin

Aruni Prabhashwari Senaratne

► **To cite this version:**

Aruni Prabhashwari Senaratne. Formation of the CenH3-independent holocentromere in Lepidoptera avoids active chromatin. Cellular Biology. Sorbonne Université, 2020. English. NNT : 2020SORUS355 . tel-03369756

HAL Id: tel-03369756

<https://theses.hal.science/tel-03369756v1>

Submitted on 7 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université

Doctoral School “Life Science Complexity”- ED515

Lab “Evolution of centromeres and chromosome segregation”

UMR3664 Nuclear Dynamics Unit, Institut Curie- Centre de Recherche

Formation of the CenH3-independent holocentromere in Lepidoptera avoids active chromatin

By: Aruni Prabhashwari Senaratne

PhD Dissertation

Directed by: Dr. Ines Anna Drinnenberg

Presented and defended publicly on 15th September 2020

Jury Members:

Dr. Ines Anna Drinnenberg- Thesis director

Dr. Reto Gassmann- Rapporteur

Prof. Elena Giulotto- Rapporteur/Examiner

Dr. Lionel Bénard- Examiner

Dr. Christophe Escudé- Examiner

Abstract

Formation of the CenH3-independent holocentromere in Lepidoptera avoids active chromatin

The centromere is an essential chromosomal locus that mediates the accurate partitioning of the genome to daughter cells during cell division. Centromeric chromatin is distinct wherein the highly conserved centromere-specific histone 3 (CenH3) (first identified as centromere protein A (CENP-A)) is the main epigenetic marker of centromeres and has been detected in the great majority of eukaryotes studied varying from protists to animals and plants. Despite the highly conserved function and composition, centromere architectures are diverse among eukaryotes and embody two main configurations: mono- and holocentromeres, referring respectively to a localized or unrestricted distribution of centromeric activity. Previous studies revealed that holocentricity in many insects strongly coincides with the loss of the otherwise essential centromere marker CenH3, suggesting a molecular link between the two events. This thesis aims to understand the molecular architecture of CenH3-deficient holocentromeres in insects by exploiting the experimental tractability of *Bombyx mori* (domestic silkworm), which belongs to the insect order Lepidoptera- one of several insect lineages in which CenH3 is lost. The work of this thesis has leveraged recently-identified centromere components in *B. mori* as well as molecular tools developed in our lab such as specific antibodies and classical genomics approaches to map the locations of centromere sites along *B. mori* chromosomes. Subsequently, we characterized the DNA sequences and chromatin elements underlying newly mapped centromere sites along *B. mori* chromosomes. This uncovered a robust correlation between *B. mori* centromere profiles and regions of low chromatin dynamics found anywhere along the chromosome and indicated *B. mori* centromere profiles to be shaped by chromatin activity. To test this hypothesis, transcriptional perturbation experiments were subsequently done to demonstrate that centromeres become excluded from active chromatin regions but can form de novo in regions where chromatin activity is low. Taken together, the identified link to chromatin dynamics allows us to discuss the plasticity of centromere identity. In this context, we present a novel mechanism of centromere formation that occurs in a manner recessive to the chromosome-wide chromatin landscape rather than being defined by the otherwise essential presence of CenH3. Based on similar profiles observed in additional Lepidoptera, we propose an evolutionarily conserved mechanism that underlies the establishment of holocentromeres through loss of a specified centromere.

Acknowledgements

A heart full of gratitude is impossible to put into words.

I am indebted to my mentor, Dr. Ines Anna Drinnenberg. Through rough beginnings and every day thereafter, thank you for your immense support and guidance. Thank you for always having your door open and for instilling in me scientific and personal values.

I thank my past and present team members (Jon, Héloïse, Nuria, Gaetan, Jose) for all your support and enthusiasm. Thank you for creating an incredibly welcoming atmosphere in the lab. I would especially like to thank Héloïse and Nuria with whom I have shared these past four years and watched the lab grow into the incredible team that it is today. Thank you for your words of encouragement and advice through and through and above all, your great sense of humor.

I thank all members of the Nuclear Dynamics Unit (UMR3664), my thesis committee members as well as all members of the Fachinetti lab for your continuous encouragement and the many stimulating scientific discussions.

I thank my family and my loved ones.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
List of Abbreviations.....	v
A note on silkmoths.....	viii
Introduction.....	1
Diversity of centromere architectures.....	3
The holocentromere: history and characteristics.....	6
Where did holocentromeres come from?.....	9
Known holocentric architectures.....	13
(I) Nematodes.....	13
(II) Plants.....	17
An Unorthodox holocentromere variant in insects.....	19
Overview of holocentricity in Insects.....	20
The role of CenH3 at centromeres.....	22
What we know about the kinetochore of CenH3-deficient holocentric insects.....	25
Research aims.....	27
Results.....	28
Result 1: The kinetochore forms a broad localization pattern along <i>B. mori</i> mitotic chromosomes.....	29
Result 2: Half of the <i>B. mori</i> genome is permissive for kinetochore assembly.....	29
Result 3: Kinetochore attachment sites in <i>B. mori</i> are anti-correlated with actively transcribed chromatin.....	32
Result 4 (i): Hormone-induced perturbations of gene expression result in proximal CENP-T loss or gain.....	34
Result 4 (ii): Differentially expressed orthologous genes in Lepidoptera show opposite patterns of CENP-T localization.....	37
Discussion.....	40
Holocentromere regulation further emphasizes centromere plasticity.....	42
Adding it all up.....	42
Evolutionary establishment of holocentric chromosomes in insects.....	45
An intrinsic potential to be holocentric?.....	47
Methods.....	50
Supplementary Figures.....	59
Bibliography.....	67
Annex.....	92

List of Abbreviations

2

2-D (two-dimensions)

3

3C (chromosome conformation capture)

3-D (three-dimensions)

20

20E (20-Hydroxyecdysone)

A

ATAC-seq (assay for transposase-accessible chromatin using sequencing)

B

blastp (protein-protein basic local alignment search tool)

bp (base pair(s))

BPM (bins per million mapped reads)

BSA (bovine serum albumin)

C

C-banding (constitutive heterochromatin or centromere-banding)

C-terminus (carboxyl terminus)

CaCl₂ (calcium chloride)

CAL1 (chromosome alignment defect 1)

CATCH-IT (covalent attachment of tagged histones to capture and identify turnover)

CCAN (constitutive centromere associated network)

cDNA (complementary DNA)

CenH3 (centromere specific histone 3)

CENP (centromere protein)

ChIP (chromatin immunoprecipitation)

ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequencing)

CPAR-1 (centromere protein A related 1)

CRM (centromeric retrotransposons of maize)

CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9)

CRR (centromeric retrotransposons of rice)

D

DAPI (4',6-diamidino-2-phenylindole)

DMSO (dimethyl sulfoxide)

DNA (deoxyribonucleic acid)

E

EDTA (ethylenediaminetetraacetic acid)

EGTA (ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid)

EM (electron microscopy)

ES (embryonic stem)

EtOH (ethanol)

F

FACT (facilitates chromatin transcription)

FBS (fetal bovine serum)	IGV (integrative genomics viewer)
FISH (fluorescence in situ hybridization)	IP (immunoprecipitation)
FLAG (peptide sequence DYKDDDDK (where D=aspartic acid, Y=tyrosine, and K=lysine))	K
	kb (kilo base(s))
	KNL-2 (kinetochore null 2)
G	L
g (units of gravity)	LCS1 (<i>Luzula</i> centromeric sequence 1)
	LTR (long terminal repeat)
H	M
H3 (histone 3)	M (molar)
H3.3 (histone 3.3)	mb (mega base(s))
H3K4me3 (tri-methylated Lysine 4 on histone H3)	Mis18BP1 (Mis18 binding protein 1)
H3K9me2 (di-methylated Lysine 9 on histone H3)	ml (milliliter(s))
H3K9me3 (tri-methylated Lysine 9 on histone H3)	mM (millimolar)
H3K27me3 (tri-methylated Lysine 27 on histone H3)	MeOH (methanol)
H3K36me3 (tri-methylated Lysine 36 on histone H3)	mRNA-seq (messenger ribonucleic acid-sequencing)
HCP-3 (histone H3-like centromeric protein 3)	N
HCP-6 (holocentric protien 6)	N-ChIP (native chromatin immunoprecipitation)
HCTD1 (Holliday junction recognition protein carboxy terminal domain 1)	N-terminus (amino terminus)
HFD (histone fold domain)	NaCl (sodium chloride)
HIRA (histone regulator A)	NaOAc (sodium acetate)
HJURP (Holliday junction recognition protein)	P
HOR (higher order repeat)	PBS (phosphate buffered saline)
	PCR (polymerase chain reaction)
	PFA (paraformaldehyde)
	pg (picogram(s))
I	R
IF (immunofluorescence)	RCS2 (rice centromeric sequence 2)
IgG (immunoglobulin G)	

A note on silkmoths.

Bombyx mori is also known as the domesticated silkmoth, resonating a long history of adaptation to human exploitation in the silk industry (Figure i1). Attempts to trace its origins began in 1926 and were led by Chinese archaeologist and lecturer at Qinghua University in Beijing, Li Ji, who recovered a spinning wheel and half of a fossilized silkworm cocoon shell from the ancient ruins of the Xiyin village in China. These remains were unearthed from remote areas of Yellow River Valley, which, intriguingly, legend has was home to the mythical Empress from the 26th Century B.C., Lei Zu (wife to China's first of five prehistoric rulers, the Yellow Emperor, Huangdi), who is believed to have initiated silkworm breeding and growing of Mulberry trees (the primary source of food for silkworms) in China. Artificial cuts on the recovered cocoon shell and man-made tools found at the excavation site allowed scientists to trace silkworm domestication and the start of silk farming (sericulture) in ancient China to at least 5000 years ago (Wei et al., 2012). Not surprisingly, phylogenetic studies indicate the direct divergence of *B. mori* from an ancestor of the Chinese wild silkworm, *B. mandarina* (Wei et al., 2012).

As opposed to its wild counterpart, *B. mori* has evidently risen across continents as an industrial workhorse that has been genetically reinvented to manufacture the world's finest quality silk, known as Mulberry silk. But also, as a tractable biological model system for studies and applications in basic genetics and engineering, biotechnology, endocrinology, and insect development and physiology. In fact, the first evidence of Mendelian inheritance in animals was demonstrated in *B. mori* by early geneticist and agricultural scientist Toyama Kametaro (Toyama 1906-cited from (Sahara et al., 2016)). Toyama's cross-breeding experiments in silkworms were the first attempts at heterosis (or heterozygosis, referring to the breeding of hybrid offspring from selected strains), a technique that is extensively used today in agriculture and farming. The resulting sturdy worms yielding improved grades of silk are regarded as Toyama's efforts to help increase the silk yields in Japan during the early 1900's (Sahara et al., 2016).

Some of the very first observations of silkworm chromosomes were also made by Toyama in as early as 1894 (Toyama 1894- cited from (Sahara et al., 2016)). Over 80 years later, in 1979, *B. mori* became the first Lepidopteran (the insect order comprising moths and butterflies) to have a correctly published mitotic karyotype (n=28) (Sahara et al., 2016). Arguably, detailed studies into the morphology of *B. mori* chromosomes began around the same time. In 1974, Murakami and Imai presented strong cytogenetic evidence supporting the holocentric nature of this organism's chromosomes (Murakami & Imai, 1974), depicting also in their microscopy images the characteristically small, numerous and often round or bar-

shaped structure of Lepidopteran chromosomes. However, the holocentric chromosomes of *B. mori* or any other Lepidopteran has until now not been investigated in detail using any high-throughput technique.



Why were silkworms in particular amenable to domestication over other silk-producing insects? In fact, spiders produce a superior silk fiber than silkworms and there have been modern attempts to domesticate spiders. However, these have failed on a large scale owing to spiders being highly territorial creatures and often resorting to cannibalism when forced to live together in an enclosure. Additionally, in contrast to harmless silkworms, the aggressive nature of spiders when confronted, such as during feeding, made it non-practical to adapt them for maximized silk production facilities, which were often the homes of Japanese families. While spiders make silk as adults for predation, silkworms make silk during larval stages, which may explain why the former are not known to be social animals, whereas the latter are even capable of co-evolving with humans that led to the human-animal partnership we refer to today as “domestication”.

Figure i1: A woodblock print from around 1800 showing Japanese women spreading mulberry leaves over a tray of silkworms. Adapted from (LeCain, 2019)

B. mori as a model system of the postgenomic era was greatly advanced by successful initiatives to sequence its genome, which led to the release of the first draft sequences in 2004 (Mita et al., 2004; Xia et al., 2004) followed by a low-quality chromosome-level assembly in 2008 (Silkworm consortium, 2008). These assemblies greatly facilitated investigations into basic silkworm biology that required genome information; for example, functional analyses by mutation of mating behaviors and disease resistance; elucidation of sex determination and small RNA (ribonucleic acid) regulation. Using long and short read sequencing technologies, a high-quality genome assembly was released in 2017, along with an accurate set of gene models (Kawamoto et al., 2019). The genome size of *B. mori* is approximately 460 mb (mega bases) (Kawamoto et al., 2019)- mid-range within the genome size limits reported in other Lepidoptera, which can range from ~200 to >1000 mb (Zhang et al., 2019). In addition to a well-characterized transcriptome and proteome (Lu et al., 2019), cell lines, as well as the realization of modern techniques to dissect the biology of this organism such as CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9)-mediated genome-editing (Ma et al., 2015) and RNAi (RNA interference) (Cortes-Silva et al., 2020; Mon et al., 2017), DNA (deoxyribonucleic acid)-FISH (fluorescence in situ hybridization) and 3C (chromosome conformation capture)-derived technologies like Hi-C (our lab), *B. mori* is becoming increasingly amenable to a wider spectrum of molecular approaches and integration into comparative genomics studies and homology analyses exploring

among other things the evolutionary history and phylogenetic relationships to an extent that was not so long ago beyond reach.

With the growth in available genome assemblies for additional Lepidopterans¹ (Ahola et al., 2014; Drinnenberg et al., 2014; Fu et al., 2018; Zhang et al., 2019), phylogeneticists have only recently begun to tap into the diversity lying within this insect order. Moths and butterflies comprise a mere 10% of the Earth's biodiversity (Mora et al., 2011) and are one of the most speciose groups of organisms on Earth. The domestic silkworm, while may not be particularly popular for its size, looks ... nor speed (to name a few), belongs to a superfamily, the Bombycoidea, which boasts over 6000 known species covering some of the largest, fastest and most attractive insect species in the world. Even among very closely-related butterflies, stark variations in genome size can be seen (Zhang et al., 2019). If anything, these differences highlight rapid diversification rates in this insect order and naturally, directs curiosity toward the most primordially-associated cue, the karyotype.

Lepidoptera have distinctly high rates of chromosome evolution- at least during more recent evolution, as reflected in substantially variable chromosome numbers across species that range from as low as $n=5$ to as high as $n>200$ (with the most common chromosome number being $n=29-31$) (Ahola et al., 2014). Yet, the total inter-species DNA content is found to be very similar, suggesting that different karyotypes have arisen through fission and fusion events of a similar number of ancestral chromosomes (Ahola et al., 2014). The feasibility to undergo such structural changes might be affiliated to the holocentric chromosomes of Lepidoptera, wherein microtubule attachments forming chromosome-wide is presumed to allow flexibility in the exact centromeric loci, such as in the event of breakage (fission) or fusion. In fact, the *B. mori* genome is thought to consist of three fusion events comprising its shortest chromosomes with six ancestral ones. Because the shorter chromosomes have more repeats- specifically being retrotransposon-enriched at fusion sites, this suggests that shorter *B. mori* chromosomes are more unstable and tend to undergo inter- and intrachromosomal rearrangements with more ease (Ahola et al., 2014)- a feature that likely extends to the TE (transposable element)-comprised genomes of other Lepidopterans as well (Ahola et al., 2014; Talla et al., 2017). Nevertheless, the species richness in this insect order is also attributed to a great number of other factors acting as driving forces in Lepidopteran evolution, including the chemical range of food plants; a variety of plant microhabitats that can be exploited for defense against predation and parasitism; as well as abiotic factors like climate change, variation in forest structures and river dynamics that influence the creation and maintenance of habitat species diversity (Solis & Pogue, 1999).

¹ A complete genome assembly for another Lepidopteran of interest to us, the cabbage looper (a.k.a the owlet moth, *Trichoplusia ni*) was released in 2018 (Fu et al., 2018). We use this genome assembly in the later part of our study for comparative experiments with *B. mori*.

B. mori continues to provide to one of the world's most sought-after finer commodities, as well as to further knowledge on insect development, endocrinology and physiology, among others. Centromere enthusiasts like us on the other hand, harness its genomic and proteomic tractability to understand a currently less-characterized phenomenon in the field- holocentricity- in all its facets, including (I) its origins, (II) its impacts on protein evolution; and (III) its impacts on centromere and chromatin organization in 2-D (two dimensions), and (IV) in 3-D (three dimensions), respectively.

In this thesis, I will describe the insights gained throughout the journey of my PhD to the first (I) and third (III) of these four aspects.

Introduction

The eukaryotic family tree shares the distinct property of having a nucleus per cell to contain that cell's genetic material. During cell division, (Flemming, Walther, 1882), it is vital that the genetic material compacted into a set of chromosomes is accurately partitioned into new daughter cells such that the genetic identity of the mother cell can be propagated throughout the life of an organism. This transmission process is mediated by an essential chromosomal structure known as the centromere. The centromere defines a specific site on each chromosome onto which assembles a multi-protein complex known as the kinetochore (Fukagawa & Earnshaw, 2014), which connects the DNA of paired sister chromosomes to the spindle apparatus in order to provide the mechanical forces needed to drive them apart (Musacchio & Desai, 2017).

Even a single chromosome that is unlatched from the spindle apparatus at the time of cell division will doom that chromosome to be rapidly lost in the cell cycles that follow. Centromeres are therefore vital for providing mitotic and meiotic stability to chromosomes. Therefore, not surprisingly, anomalous centromere function culminates in a defective genome with an unusual karyotype (aneuploidy). As detrimental as the phrase sounds, it is so. Briefly, if not lethal at the embryonic stage, aneuploidy during meiosis can seriously reduce the quality of life and is associated with profound medical disorders in new-borns, such as the most common human aneuploidy, Down syndrome. Centromere defects in the form of dicentric chromosomes (chromosomes with two functional centromeres), have also been linked to genomic rearrangements and human tumorigenesis (Gascoigne & Cheeseman, 2013) as well as Turner's syndrome, where 15% of all patients exhibit stable dicentricity (Chapelle et al., 1966; McNulty & Sullivan, 2017; Ockey et al., 1966). In light of critical function, the centromere is thus painted as a highly unadaptable structure that may resist change in order to maintain fine-tuned control over the chromosome segregation process.

This is however not the case. In stark contrast to its highly conserved function, the centromere architecture or organization at the molecular level, is in itself astonishingly diverse among eukaryotes. On the one hand, simple eukaryotes like most fungi to the more complex multicellular organisms like mammals and most plants share an overall centromere organization that comprises a central "centro"chromatin domain containing the centromere-specific histone known as CenH3 (centromeric specific histone 3) (first described as CENP-A (centromere protein A) in humans) (Earnshaw & Rothfield, 1985; Palmer et al., 1991) that is in turn flanked along its perimeter by highly compact blocks of chromatin enriched in silent histone marks-collectively referred to as "peri"centric heterochromatin (S. W. Brown, 1966; Ho et al., 2014; Sullivan & Karpen, 2004). Yet, this common hc-cen-hc (heterochromatin-centromere-heterochromatin) organization hugely under-represents

the great variety in overall size, underlying DNA sequence, and chromatin signatures found at the centromere regions of individual species of these lineages. On the other hand, an abundance of eukaryotes completely lacking the basic hc-cen-hc organization are also present.

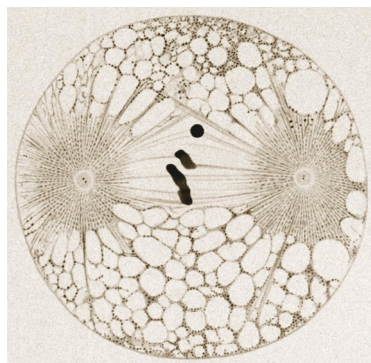
In the following sections, I will outline the diverse types of centromere architectures that are found among eukaryotes, with a specific focus on the centromere type known as the holocentromere. In the context of holocentromeres, I will discuss the molecular architecture of a handful of holocentromere variants that have been profiled to date in several model systems. This will lead me to summarize our previous work in which we identified a unique holocentromere variant akin to certain insect lineages in that these insects lacked the highly conserved centromere-specific protein CenH3. In the known context of the importance of CenH3 for centromere function, I will then outline key questions which arose from this discovery about the basic centromere biology of CenH3-deficient holocentric insects. This will bring me to the focus of my PhD project, which is to understand the molecular architecture of CenH3-deficient insect holocentromeres. My results will finally lead me to discuss an alternative, under-weighed model for CenH3-independent centromere regulation. In the context of this model, I will share the overall insights we obtained from this study into holocentromere evolution.

Diversity of centromere architectures

Walther Fleming was an anatomist who discovered chromosomes and correctly described their sequential movement during mitosis starting from the mother nucleus to the point of inclusion into daughter nuclei. When observing his stained preparations of salamander gills and fins under his microscope, Fleming noticed that the chromosomes, or “mitosen” as he called them, had a single region onto which spindle fibers attached. He referred to this region as the primary constriction (Fleming, Walther, 1882), that we synonymously refer to today as the centromere (Darlington, 1936). Only six years later, embryologist and trained artist, Theodor Boveri (Figure i2, right), was studying the process of meiosis in *Ascaris* embryos together with his wife and biologist Marcella O’Grady, when he noticed that the chromosomes of this parasitic nematode lacked a primary constriction. In his illustrations (Figure i2, left), Boveri clearly depicted that, as opposed to Fleming’s observations, the spindle fibre attachments extended along the entire surface of *Ascaris* chromosomes (Boveri, Theodor, 1888). These two independent observations from the late 19th century can collectively be considered the first evidence that centromeres can indeed exhibit different morphological forms. It was however only thirty-five years later that this difference came to be acknowledged and the latter type was termed as the holocentromere (or holokinetochore) by Franz Schrader (Schrader, Franz, 1935). At the very basic level, this difference in microtubule attachment served as a hallmark that allowed early cytologists to differentiate between the two most common centromere architectures that are known to us today, with the former type being known as the monocentromere. This criteria relating to the presence or absence of a primary constriction and spindle distribution pattern contributed to much of the early literature on the suspected mono-/holocentricity of several eukaryotic lineages (discussed later).

A century after Fleming described centromeres using his innovative microscopy techniques, the first centromere was characterized at the molecular level. This was that of the budding or baker’s yeast *Saccharomyces cerevisiae* (Clarke & Carbon, 1980). Progressive work starting in 1982 and spanning over twenty-five years has unravelled the budding yeast centromere to have a distinct tripartite DNA sequence motif of ~125 bp (base pairs) that is essential for centromere function, which specifically incorporates a single CenH3-containing nucleosome and binds a single microtubule (Fitzgerald-Hayes et al., 1982; Furuyama & Biggins, 2007). This organization is referred to as a point centromere, reflecting the 1:1 stoichiometry in centromere components within a small restricted site. The point centromere is the simplest monocentromere known to us and is akin to a handful of other budding yeasts of the same sub-phylum (saccharomycotina) (Dujon et al., 2004; Meraldi et al.,

2006); although in some, the highly-conserved (consensus) tripartite DNA motif appears to be entirely different (N. Kobayashi et al., 2015).



(Satzinger, Helga, 2008)



(Groeben, 2019)

Figure i2: An instance where Boveri flaunted his highly trained hand at drawing can be seen on the left, which shows a lithograph of an *Ascaris megalocephala* cell in division, 1901. The characteristically large *Ascaris* chromosomes lacking a primary constriction and with spindle microtubules attached along the polar length of sister chromatids were visualized by Boveri using early light microscopy and staining techniques of the day. On the right, Boveri can be seen in the distance enjoying a day out in Naples in the spring of 1896. Boveri visited the Naples station from time to time, which was the first urban-placed research institute in marine biology. Here, Boveri collected sea-urchin specimens from the Mediterranean Sea in order to study chromosome inheritance and embryonic development in these organisms. This laid the groundwork for his infamous chromosome theory of inheritance in collaboration with geneticist and physician, Walter Sutton.

The genetically defined point monocentromeres of budding yeasts are the exception to the rule because the majority of monocentric organisms have regional monocentromeres that are defined in a sequence-independent manner (discussed later). As opposed to budding yeasts, regional monocentromeres are contained within a much larger span of the chromosome, at times exceeding several mb in length (McKinley & Cheeseman, 2016; Muller et al., 2019), hence deriving the name. Regional monocentromeres ranging from ~4-300 kb (kilo bases) are found in fungi, including in the second most widely studied ascomycete *S. pombe* (fission yeast) and the pathogenic yeast *Candida albicans*. Collectively, the diversity in the features of regional monocentromeres found in the fungal kingdom can to some extent be consolidated by the lack of conserved recognition sequences for targeting the essential centromere component CenH3², presence of multiple CenH3-containing centromeric nucleosomes attaching to multiple microtubules, well-defined pericentric heterochromatin enriched for silent histone marks like H3k9me2/3 (di/tri-methylated

² In the centromeres of the first basal fungus to ever be characterized, which belongs to the Mucormycotina, CenH3 is unidentified. However, it also has hybrid features of both a point and regional monocentromere, including a centromere-specific DNA sequence motif (Navarro-Mendoza et al., 2019), therefore blurring its classification into one or the other category.

Lysine 9 on histone H3)³, and centromeric or centromere-flanking repeats⁴ (reviewed in (Friedman & Freitag, 2017)) -key characteristics that extend to the regional centromeres of animals and plants.

Regional monocentromeres of animals and plants can span up to several mb in size (Steiner & Henikoff, 2015), where in particular, the presence of underlying repeat sequences is a prominent feature (Melters et al., 2013; Steiner & Henikoff, 2015). Distinct classes of centromeric repeats are affiliated to the centromeres of different organisms. For example, the regional centromeres of monocentric plants such as *Arabidopsis thaliana*, rice and maize are characterized by the presence of simple tandem repeats (also known as satellite DNAs) and retrotransposon families that are unique to each species (e.g. Athlia and LTR (Long Terminal Repeat) retrotransposons in *Arabidopsis*; CRM (centromeric retrotransposons of maize); and CRR (centromeric retrotransposons of rice) (reviewed in (Muller et al., 2019)). In mice, functional centromeres form on 120 bp satellites known as minor satellites, whereas in human (and all primates), satellite monomers known as alphoids of 171 bp length (Waye & Willard, 1987) come together to form HOR (higher order repeat) structures (reviewed in (Muller et al., 2019)). Furthermore, a conserved 17 bp DNA motif known as the CENP-B box is found embedded within the centromeric satellites of mammalian centromeres (Masumoto et al., 2004). In contrast, some monocentric organisms have naturally repeat-less centromeres- first discovered in the satellite-less centromere of chromosome 11 of the horse (Wade et al., 2009); and thereafter multiple chromosomes with satellite-less centromeres in other equids (Nergadze et al., 2018); several tandem repeat-less centromeres in chicken (Shang et al., 2010); the satellite-less centromere on chromosome 12 of orangutan (Locke et al., 2011); as well as several satellite-less centromeres in potato (Gong et al., 2012).

At the opposite spectrum of this variety of architectures that are collectively grouped as variants of the monocentromere, lies the holocentromere. The holocentromere, as mentioned earlier, is characterized by centromere activity that rests on extensive parts or even the entire polar surface of the chromosome. This architecture is of particular interest to our lab for reasons that emerge later (outlined in Introduction), which, also served as the basis of my PhD research. Thus, is explored in detail in the sections that follow.

3 One naturally occurring fission yeast strain lacks pericentric heterochromatin (W. R. A. Brown et al., 2014).

4 The regional centromeres of *C. albicans* lack pericentric repeats in the majority of chromosomes. However, inverted repeats of considerable length (~ 520 bp to ~4.8 kb) flanking the centromere core are found on two of its chromosomes (Sanyal et al., 2004).

The holocentromere: history and characteristics

Holocentric chromosomes were recurrently documented since Boveri's first illustrations in 1888, although never actually characterized per se until over thirty-five years later (Schrader, Franz, 1935). Starting in 1935, and throughout decades of collaborative work that followed, Franz Schrader with his wife and zoologist, Sally Hughes-Schrader (Figure i3) used cytogenetic techniques and induced stresses on Hemipteran insects (which includes true bugs, cicadas, hoppers, aphids and allies) to demonstrate a number of characteristics (additional to the emblematic lack of a primary constriction) (Figure i4) that are unique to this centromere architecture (Hughes-Schrader & Schrader, 1961; Hughes-Schrader, Sally, 1944; Schrader, 1947; Schrader, Franz, 1935).



Figure i3: Franz (Left) and Sally (Middle) Schrader and Mary R. Huettner (Right) at Gansett Beach, 1923. When not collecting insect specimens or studying their cytological behaviors, the Schraders enjoyed active pursuits such as swimming and boating around Woods Hole, Massachusetts. Here, they can be seen sitting on a piece of the wrecked hull of the Wanderer, a former whaling ship that crashed on the rocks at Cuttyhunk in 1921. (Adapted from Marine Biology Laboratory Archives).

The basic criteria that were shaped by their observations and are still used today for preliminary identification of holocentric chromosomes are outlined below:

- 1) Diffuse kinetochores cover the polar length of mitotic sister chromatids and serve as the basis for spindle fibre attachments along the entire poleward surface of chromosomes.
- 2) As a result of chromosome-wide spindle fiber attachments, sister chromatids exhibit parallel disjunction at anaphase (as opposed to the “V”-shaped movement of monocentric chromosomes where the chromosome arms lag behind the centromere).

3) Due to the presence of scattered spindle attachment points, chromosomal fragments (for example, from X-ray induced breakage) will retain kinetic activity even after being separated from the original whole chromosome. These fragments are mitotically potent over many cell divisions.

4) At meiosis, specific structural adaptations are present⁵: The Schraders noted that in spermatocyte chromosomes of stink bugs (Sub-order: Heteroptera), there is a temporary restriction of kinetic activity, where the diffuse spindle seen at mitosis is altered such that it is now confined to a terminal region of the chromosome, thus temporarily behaving like a monocentric chromosome. Additionally, when studying the chromosome behaviors in some coccid species (scale insects) collected in Mexico, Sally Hughes-Schrader noted the apparent loss of all cohesion on sister chromatids in meiosis I (Hughes-Schrader, Sally, 1944)- a feature that is usually not observed until meiosis II, as described in footnote #5. This phenomenon is now referred to as “inverted meiosis⁶”. These observations are examples of meiotic adaptations in

⁵ Meiotic adaptation is necessary for holocentric chromosomes because of the complications they would otherwise face due to the diffuse nature of their microtubule attachments. Typically, at meiosis, following genetic exchange between homologous chromosomes (homologous recombination), chromosomes must separate from each other (meiosis I, or reduction division). In monocentric chromosomes, this separation is facilitated by the selective retention of cohesion only at the single centromere region, while cohesion is “released” from the centromere-distal regions. This allows sister chromatids of individual chromosomes (held together at the centromere by the cohesin protein complex) to remain together until their partition into gametes at much later stages (meiosis II) (reviewed in (Dernburg, 2001). However, for a holocentric chromosome, because there is no defined centromere locus, the regions at which cohesion must be released or maintained at meiosis I are obscure, therefore blurring the distinction between a homologous chromosome and a sister chromatid. Therefore, the question arises as to how in this case, sister chromatids ensure that they maintain cohesion until meiosis II, while homologous chromosomes can release cohesion and segregate at meiosis I? (Dernburg, 2001). Holocentric organisms have evolved specific structural adaptations with regard to how the spindle attaches along the chromosome so that correct, bi-oriented separation of chromosomes can take place. One of these adaptations is to limit the kinetochore activity to the chromosomal ends such that pulling forces are exerted in only two opposite directions. In addition to the Schraders’ observations in the Hemiptera, restricted kinetochore activity is also described in holocentric nematodes like *Caenorhabditis elegans* (Melters et al., 2012) and *Parascaris univalens* (Goday & Pimpinelli, 1989) as well as several other hemipterans such as *Triatoma infestans* (kissing bugs/barber bugs) and milkweed bugs (Melters et al., 2012).

⁶ Inverted meiosis happens when the steps of meiosis I and II are reversed. In typical meiosis, homologous chromosomes will first separate from each other (reduction division) (meiosis I). Then, in meiosis II, sister chromatids of each chromosome will separate (similar to mitosis) to form four haploid gametes. In inverted meiosis, sister chromatids of each homologous chromosome will first separate in meiosis I. Then homologous sister chromatids re-pair prior to meiosis II to reform a “chromosome” that can then be partitioned into gametes (meiosis II). This switch in the steps allows diffuse kinetochore activity to be maintained. In addition to coccids, inverted meiosis is also reported in other animals and plants including some dragonflies and damselflies, spiders, and woodrushes (Melters et al., 2012).

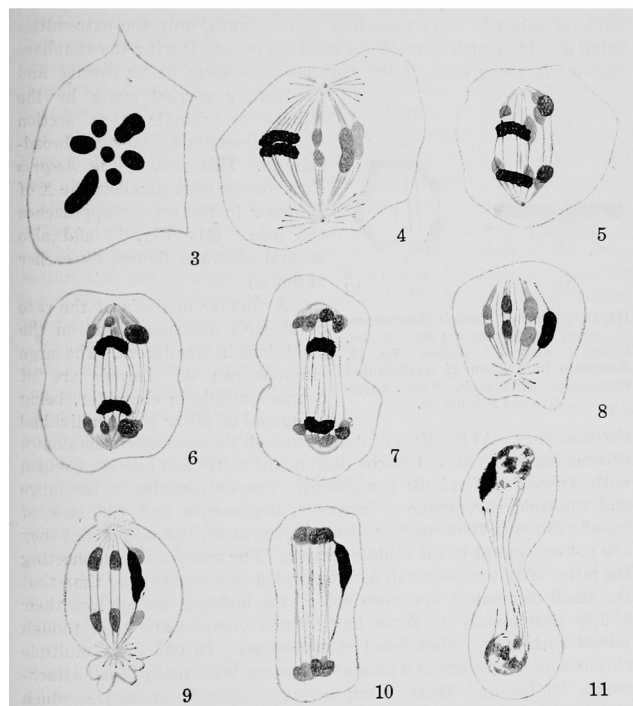
holocentric organisms, the molecular mechanisms of which are currently not well understood.

F. Schrader further speculated on two possible types of holocentric architectures that could reconcile the broad nature of spindle fibre attachments discernible to him by eye. His interpretations referred to two different patterns of kinetochore distribution along the chromosome (Schrader, 1947; Steiner & Henikoff, 2014):

1) The polycentric model (also known as the multiple kinetochore): the holocentromere is composed of a chain of individual kinetochore units spread along sister chromatids, each of which can connect to a spindle fiber.

2) The diffuse model (also known as the diffuse kinetochore): the holocentromere is composed of kinetochores truly distributed chromosome-wide, where spindle fibers attach randomly.

Either way, both models support an architecture that would impart similar kinetics to mitotic chromosomes.



(Schrader, Franz, 1935)

Figure i4: F. Schrader's drawings of dividing *Protenor* chromosomes from spermatocyte cells as seen by light microscopy techniques of the day. Characteristics such as the lack of a primary constriction, chromosome-wide microtubule attachments, and parallel migration of sister chromatids can clearly be seen. 15 X.

Where did holocentromeres come from?

It must now be apparent that scientists were very aware of the fact that centromeres of different organisms are indeed not conformed to a single structure. There is no doubt today that holocentric organisms are widespread across diverse eukaryotic lineages including nematodes such as the well-known *C. elegans* to a range of arthropods and flowering plants (Melters et al., 2012) as well as zygnematophyceae algae, the only known holocentric class of charophyte algae (Zedek & Bureš, 2018). As with any other attempt to explain the origins of life, it can only be speculated where lies the beginnings of holocentricity. Phylogenetic analyses to trace evolutionary relationships is particularly advantageous in the aspect that it allows the simultaneous comparison of hundreds to thousands of subjects, and importantly, to assimilate patterns in the distribution of a certain trait in question that we would otherwise remain oblivious to at the individual scale. What *has* been repeatedly revealed in such independent studies that have compared large groups of both mono- and holocentric organisms is that the trait of holocentricity has a sporadic pattern of distribution in the eukaryotic family tree, therefore making it at least apparent that this was an adaptive trait that was acquired by only some lineages and not others (also called adaptive evolution or convergent evolution), and even still, by some species and not others of the same family (Melters et al., 2012). An overarching pattern where entire holocentric clades are found positioned nested within larger monocentric groups (Drinnenberg et al., 2014; Escudero et al., 2016; Melters et al., 2012), has shaped the current consensus in the field that the holocentric trait was derived from monocentric ancestors. The lack of conclusive evidence pointing to reversions to the monocentric form in any eukaryotic lineage further supports the unidirectional evolution of holocentricity and suggests it to be a stable trait once adapted. This idea that a monocentric form preceded the holocentric one is in fact apparent in the numerous models and evolutionary drivers that have been put forward over the years based at times on single-species studies, which, repeatedly conclude a monocentric ancestral state. I briefly outline these models below:

1) Centromere drive suppression (Malik & Henikoff, 2009): Female meiosis is asymmetrical in animals, plants and some ciliates (Malik & Henikoff, 2009). This means that, during the formation of mature female gametes, the chromosomes in primary oocytes which get included into the mature egg are determined by the strength of their centromeres. In other words, those with weaker centromeres are eliminated as polar bodies, while those with stronger centromeres (also called “selfish centromeres”) get incorporated into the single surviving egg. Greater centromere strength is thought to be imparted to a chromosome by the expansion of the centromere region, thereby biasing the spindle strength in favor of the transmission of that particular chromosome into the egg (Akeru et al., 2017; Chmátal et al., 2014).

The expansion of centromere regions in this way is called “centromere drive” (Henikoff, 2001), and is usually interpreted as the expansion of satellite DNAs that are typically found at the regional centromeres of eukaryotes, resulting in the concomitant expansion of the CenH3-bound chromatin window (Chmátal et al., 2014; Iwata-Otsubo et al., 2017). Thus, centromeres would be in a continuous arms race with each other for a greater transmission advantage. The holocentric architecture is proposed to have evolved as a defense mechanism to prevent the harmful consequences of this battle, where the presence of centromeres chromosome-wide could impart to each chromatid the same likelihood of binding to the spindle, therefore balancing out any favoritism for spindle attachment. This model however does not offer an explanation for the evolution of holocentrism in mitosis- a feature that all currently known holocentric organisms have in common (Marques & Pedrosa-Harand, 2016; Zedek & Bureš, 2018). Furthermore, there are many lineages that are holocentric in mitosis but because of meiotic adaptations, are essentially monocentric in meiosis (e.g. Hemiptera as described in previous section) (Marques & Pedrosa-Harand, 2016) during which this battles takes place. Therefore, this model does not reconcile this plasticity of meiotic holocentromeres.

2) Kinetochore formation parallel to the chromosomal axis (Nagaki et al., 2005): Here, the primary constriction of monocentric chromosomes is proposed to be linked to the presence of the kinetochore in a position that is orthogonal to the chromosome axis. The authors proposed that if instead, the direction of kinetochore formation turns by 90 degrees so that kinetochore formation now happens along the direction of the chromosomal axes (i.e. sideways), this would give rise to holocentric chromosomes, because of the horizontally-directed assembly of kinetochore components, resulting in an extended primary constriction spanning the sister chromatid. The authors support this model by the observation that a centromere “groove” structure resembling an extensive primary constriction can be seen on mitotic chromosomes of holocentric plants (Heckmann et al., 2011; Marques et al., 2015; Nagaki et al., 2005). This model however does not explain the chromosomes of holocentric animals, where the centromere “groove” has not been observed.

3) Telomere to centromere model (Villasante et al., 2007): This model proposes that holocentromeres arose from the telomeric sequences of monocentric chromosomes, which could have spread by amplification throughout internal chromosomal regions. The authors draw support for this model from the repeat distribution pattern in the classic holocentric model system, *C. elegans*, where tandem repeats, including telomeric repeats are found at internal sites of its chromosomes with a higher abundance at chromosomal ends. Although some repeats are also found in certain holocentric plants (Haizel et al., 2005; Marques et al., 2015), repeat distribution (or presence altogether) is not characterized broadly enough among holocentric organisms to generalize this theory and as of now, only a high degree of heterogeneity

in repeat distributions among holocentric organisms can be observed (Melters et al., 2012). In addition, in *C. elegans*, which was used to support this model, centromeres are not associated with repeats (Gassmann et al., 2012).

4) Spreading of satellites (Neumann et al., 2012): Based on the observation that in the holocentric plant *Luzula nivea*, satellite DNA LCS1 (*Luzula* centromeric sequence 1) shows similarity to the monocentric rice satellite RCS2 (rice centromeric sequence 2), Neumann et. al. extended the centromere drive hypothesis in that satellite sequences present at monocentromeres may have spread chromosome-wide. The authors further draw support to this theory from pea plants, which also have multiple clusters of satellites localizing along their meta-polycentromeres (elongated regional monocentromeres containing several distinct CenH3-enriched regions). These are proposed to represent an intermediate form between the mono- and holocentric forms.

5) Nematode development through fixed cell lineages (Pimpinelli & Goday, 1989): Goday and Pimpinelli presented a different hypothesis driving holocentricity within the specific context of nematode development. Nematodes undergo “mosaic development” that gives rise to fixed somatic cell lineages. That is, each pre-cursor cell undergoes a pre-determined number of divisions to differentiate into a pre-determined terminal state that is necessary for the fitness and survival of the entire multicellular organism. Thus, in most cases, one cell cannot be compensated by, nor compensate for, another cell. Given this critical sequence of events for correct development, even a single inappropriate cell death during the developmental process could be, in the worst case, lethal to the embryo. The authors speculate that holocentricism may have been selected in nematodes as a defense against the lethal effects of DNA damage such as an un-repaired chromosome breakage that could lead to loss of acentric fragments and cell death, which in this case could critically affect the progression of development.

6) Environmental stimuli (Zedek & Bureš, 2018): Zedek and Bures have presented numerous examples from both experimental and natural (Figure 15) contexts, where holocentric species demonstrate greater resistance over monocentrics to certain clastogenic stimuli such as UV (ultraviolet), gamma and X-ray irradiation. Based on the knowledge that this trait can protect against associated chromosomal fragmentation and loss (Hughes-Schrader & Schrader, 1961), the authors argue by correlating the evolutionary time points at which holocentric clades have diverged -to the time of early terrestrial colonization, to put forward the theory that holocentricity could have been a valuable adaptation that offered resistance against exposure to various natural sources of radiation dominating the Earth’s atmosphere at the time. They draw particular support to this theory from the phylogenetic distance observed between land plants and the Zygnematophyceae- which, interestingly, are the only known holocentric lineage of algae that also happen to be the closest relative to land

plants. Here, the authors argue that holocentric chromosomes may have allowed this relatively primitive eukaryote to leave the oceans, move to a new environment and colonize land half a billion years ago by directly counteracting the clastogenic effects of the Sun's cosmic radiation on its DNA. The authors find additional examples in the feeding habits of insects, where they point out that the majority of phytophagous insects species belong to holocentric orders of the Hemiptera and Lepidoptera. This link is associated by them to the clastogenic effects of chemical compounds present in certain plants fed on by these insects (such as the presence of nicotine in tobacco on which aphids (hemiptera) feed).



(Miehe et al., 2008)

Figure i5: As opposed to a handful of experimental studies that demonstrate the resistance of holocentric organisms to clastogenic stimuli, isolated carpets of the bog sedge *Kobresia Pygmaea* found along the intensely-UV-exposed Tibetan highlands might offer a purely natural example of such endurance. *K. pygmaea* turfs reach altitudes as high as 6000 meters on the slopes of Mount Everest and dominate over a coverage between 90-98%. *K. pygmaea* belongs to the Cyperaceae, one of the two monocot holocentric plant families.

With only preliminary or marginal support of a selective advantage, the last driver (6) being perhaps the most generalizable factor driving the transition to holocentricity in distant, non-related eukaryotic lineages which in particular does not imply a prior context of the status of DNA sequence (as is the case for most models proposed). The importance of considering this criterion becomes apparent when zooming into the holocentromere architectures that have been characterized to date in different organisms (discussed below), where in a nutshell, no functional correlation between centromeres and DNA sequence can be identified.

Known holocentric architectures

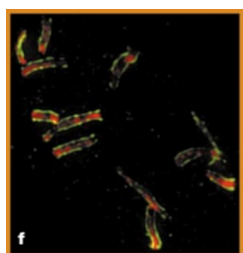
Historically, a combination of the four criteria put forth by F. Schrader (discussed previously) has been applied by early cytogeneticists to determine whether an organism possesses holocentric chromosomes, where the fate of irradiated chromosomal fragments served as the ultimate exam. However, limited often by the small size of chromosomes in certain species and poor cytological resolution of early light microscopes, to this day much remains to be verified about the inferred holocentricity in some clades. For present day holocentromere enthusiasts, much of these hurdles are overcome because of the availability of fine-tuned molecular methods to precisely confirm the presence of holocentric chromosomes in an organism. Owing to a growing number of sequenced genomes and molecular tools such as centromere- or kinetochore-specific antibodies coupled with high-resolution microscopy techniques and high-throughput sequencing technologies, the nature of occupancy of kinetochores and spindle microtubules along a chromosome can be determined with high confidence. Indeed, scientists have used such approaches - particularly immunolocalization techniques- to target and profile the centromere-specific protein CenH3, to get insights into the holocentromere architecture in a handful of organisms. This has contributed majorly to our current understanding of holocentromere organization at the molecular level. These are discussed below.

(I) Nematodes

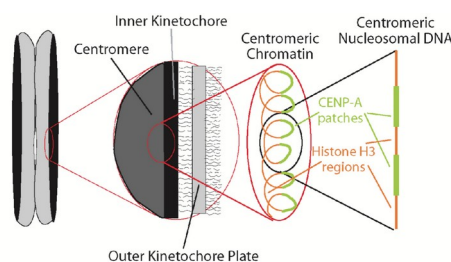
The presence of holocentric chromosomes was documented very early on in at least two parasitic nematode species: the previously described small-intestinal roundworm of the horse *A. megalocephala* (known today as *Parascaris*) by Boveri (Boveri, Theodor, 1888); and almost a century later, in the plant-parasite known as the northern root nematode (*Meloidogyne hapla*) by Goldstein and Triantaphyllou (Goldstein & Triantaphyllou, 1980). Whereas Boveri used a combination of light microscopy, advanced staining techniques of the day and his exceptional artistic skills to depict the chromosome structures in *Ascaris*, the latter group used EM (electron microscopy) techniques to visualize the diffuse kinetochores along *M. hapla* meiotic and mitotic chromosomes that they described to move with “no visible trailing of the arms” – referring to the characteristic parallel movement of holocentric sister chromatids toward opposite spindle poles.

Holocentromere organization is however best characterized in the genetically-docile model nematode *C. elegans*. Several of the earlier described characteristics associated to the holocentric architecture (i.e. lack of a primary constriction and presence of diffuse kinetochores and microtubule attachments on mitotic chromosomes as well as

the mitotic stability of radiation-induced chromosomal fragments) were first demonstrated in *C. elegans* using light microscopy and EM by Albertson and Thompson in 1982 (Albertson & Thomson, 1982), thus confirming very early on that holocentricity is a trait that extends to multiple nematode orders⁷. In 1999, Buchwitz et al. used immunolocalization techniques to show that the centromere-specific protein CenH3 (HCP-3 (histone H3-like centromeric protein 3) in *C. elegans*) localizes as longitudinal bands that span the polar length of condensed *C. elegans* chromosomes (Figure i6, Left) (Buchwitz et al., 1999). These images are the first to vividly show the diffuse kinetochore plates on holocentric chromosomes that are described in previous literature. It was proposed that this “band like” distribution of kinetochores is a result of higher order chromosomal rearrangements at mitosis that bring together in space “centromeric modules” that are otherwise scattered throughout the *C. elegans* interphase chromatin that are visible as “foci” in interphase nuclei (Albertson & Thomson, 1982; Buchwitz et al., 1999). Indeed, a general model for holocentromere organization has been put forward that describes the formation of distinct domains containing primarily CenH3 nucleosomes that could present centromeric chromatin as a unit at the polar surface of sister chromatids while packaging histone H3 nucleosomes inwards to form the chromatin bulk of the chromosome (Maddox et al., 2004). This model is in fact similar to what is speculated to be the case for monocentromere organization as well (Blower et al., 2002) and was to become a model put forward by independent groups for additional holocentromeres that were to be profiled in the future.



(Buchwitz et al., 1999)



(Maddox et al., 2004)

Figure i6: Left: IF (Immunofluorescence) image showing diffuse CenH3^{HCP-3} staining (green) along mitotic *C. elegans* chromosomes (red). Scale bar 10 μ m (micrometers). Right: Model for formation of CenH3-containing chromatin domains on a condensed holocentric chromatid. Interspersed regions of histone H3-containing nucleosomes and CenH3-containing nucleosomes are organized such that CenH3-containing domains face outward and direct formation of the outer domains of the kinetochore, while histone H3 domains are internal and constitute the chromatin bulk.

⁷ While holocentricity appears to be widespread in nematodes, some monocentric species are reported. Whereas some others are inferred to be holocentric based on the phylogenetic relationship to *C. elegans* rather than based on karyotypic evidence. Still others, like the causative agent of river blindness (*Onchocerca volvulus*) have conflicting reports for possessing holocentric chromosomes (Melters et al., 2012).

It still remained a subject of speculation though what factors must be involved in determining where CenH3 modules formed along the overall template that is the *C. elegans* chromosome. Two recent studies have profiled CenH3 localization patterns along *C. elegans* chromosomes using DNA microarray technology and high-throughput sequencing, respectively (Gassmann et al., 2012; Steiner & Henikoff, 2014). These studies revealed a wealth of information about both the DNA and chromatin signatures underlying *C. elegans* centromeres that have allowed us to infer the role of molecular components in determining how CenH3 localizes to where it does along this organism's chromosomes. At the level of primary DNA sequence, the identified CenH3-enriched regions were found to be generally repeat-poor and comprised of a heterogeneous, unique set of DNA sequences (Gassmann et al., 2012). At the chromatin level, CenH3-enriched regions were also enriched for H3K9me3-marked heterochromatin (Steiner & Henikoff, 2014). Additionally, the above discussed model by Maddox et. al. proposing interphase CenH3 nucleosomes to be interspersed among canonical H3 nucleosomes could be experimentally reconciled based on CenH3 quantifications (Gassmann et al., 2012). It was further revealed that in later embryonic stages, CenH3 assembles into low occupancy CenH3 domains which span transcriptionally-silent chromatin that are interspersed among discrete, high occupancy CenH3 sites (Figure i7) (Gassmann et al., 2012; Steiner & Henikoff, 2014). CenH3 incorporation into transcriptionally-silent chromatin was inferred to be a result of low nucleosome turnover in those regions, which delimited CenH3 to these sites by way of exclusion from regions with high nucleosome turnover. In very early, transcriptionally-inactive embryonic stages, this simple inverse relationship to transcription might however be more complex. Instead, cues transmitted from the female-germline related to its transcriptional status might determine centromeric profiles in these early embryonic stages (Gassmann et al., 2012). Additional tangible evidence using photobleaching experiments demonstrated CenH3 to not be stably inherited in dividing *C. elegans* embryos (Gassmann et al., 2012). Notably, these collective results brought to attention that as opposed to the general consensus that the majority of eukaryotic centromeres need epigenetic cues related to pre-existing CenH3 to propagate centromere identity (discussed later), centromere patterns could indeed be shaped independently of any prior memory of CenH3's location from the mother cell- as was evident in the complete turn-over of this protein between embryonic divisions. We will later revisit these observations in the broader context of the subject matter of this thesis.

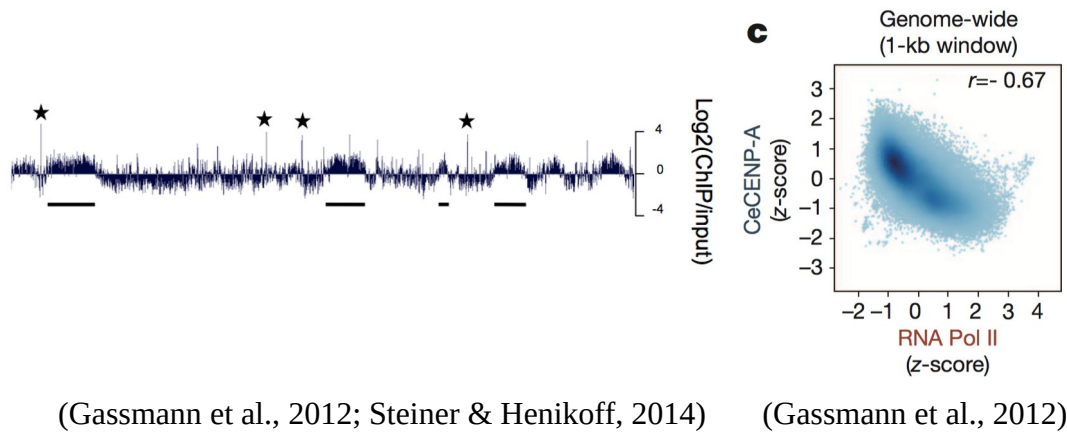


Figure i7: Left: Genome browser snapshot showing regions enriched for *C. elegans* CenH3^{HCP-3}. Low occupancy CenH3 domains (black bars) and high occupancy CenH3 sites (black stars) are indicated. Right: Genome-wide correlation plot of *C. elegans* CenH3^{HCP-3} and RNA Pol II (RNA Polymerase II) occupancy. The correlation coefficient (r) is in the upper right corner.

(II) Plants

Holocentric chromosomes are inferred to have evolved at least four independent times in the plant kingdom (Melters et al., 2012). Specifically, this trait is confined to the flowering plants (Angiosperms) and include species of both monocots and eudicots (Melters et al., 2012). Recent groups have applied fluorescence light microscopy to visualize CenH3 staining patterns in two monocots belonging to the Juncaceae family, the snowy woodrush (*Luzula nivea*) and *Luzula elegans*. In both, CenH3 forms diffuse dot-like foci in interphase that merge during mitosis to form the characteristic continuous lines along the polar length of sister chromatids (Heckmann et al., 2011; Nagaki et al., 2005). A tandem repeat monomer of length 173-178 bp (LCS1) that was conserved in most *Luzula* species and that has homology to rice RCS2 repeats was also located. LCS1 monomers form up to 50 kb long arrays that co-localize at heterochromatic sites which are also enriched for centromere-specific proteins (Haizel et al., 2005). However, no centromere-specific role for these repeats has been described.

The holocentromere architecture is also characterized in one member of the monocot Cyperaceae family, the beak-sedge (*Rhynchospora pubera*). Using IF- FISH and CenH3 ChIP-seq (chromatin immunoprecipitation followed by high throughput sequencing), *R. pubera* holocentromeres were revealed to be enriched in 172 bp long satellite DNAs (Tyba -1 and -2 satellites) (Figure i8) that form 3-16 kb long arrays that are further organized into HORs- similar to the centromeres of most other eukaryotes. Similar to most plant monocentromeres moderate levels of LTR retrotransposons belonging to the Ty3/gypsy clade (termed as CRRh) were also found (Figure i8). While the satellite DNAs were found to be transcriptionally active, a functional role for any of the identified DNA sequences or transcripts at *R. pubera* centromeres remains unknown. Nevertheless, the presence of dot-like CenH3 foci at interphase that were observed to merge into continuous lines on mitotic chromosomes (as seen for *Luzula* species in independent studies (Heckmann et al., 2011; Nagaki et al., 2005)) led the authors to propose a model similar to that put forward by Maddox in 2004, where interspersed CenH3 nucleosomes in interphase are described to be placed at the chromosome surface upon condensation (Marques et al., 2015).

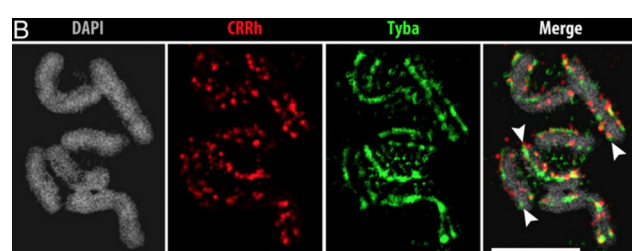


Figure i8: SIM (structured illumination microscopy) image showing DNA-FISH with retrotransposon CRRh (red) and Tyba 1+2 (green) on metaphase chromosomes of *R. pubera*. Arrowheads indicate the longitudinal centromere groove. Scale bar 5 μ m.

(Marques et al., 2015)

Recently, immunodetection techniques have also been applied to visualize the diffuse nature of microtubule binding along the chromosomes of a holocentric, parasitic eudicot, *Cuscuta* (dodders) (Oliveira et al., 2020). These were the first precise experiments to confirm holocentricity in this genus as opposed to previous reports based on chromosome morphology and mitotic and meiotic behaviors (reviewed in (Oliveira et al., 2020)).

An Unorthodox holocentromere variant in insects

While these studies have provided us a significant understanding of the spectrum of holocentromere architectures out there, our previous studies interrogating the kinetochore composition changes associated with holocentricity revealed yet another variant that lay on the extreme other end. As opposed to the above discussed types, this holocentromere type lacked the centromere-specific histone H3 variant CenH3 and with the exception of a few Hemipteran species was unique to all holocentric insect lineages tested (Cortes-Silva et al., 2020; Drinnenberg et al., 2014) (See: *Overview of holocentricity in insects*).

CenH3 and canonical H3 share a highly conserved HFD (histone fold domain) at the C (carboxyl)-terminus of their protein structure. However, in order to distinguish the two proteins, bioinformatic criteria pertaining to several unique features of CenH3 can be used. These include the presence of a longer, highly divergent N (amino)-terminal tail (as opposed to the invariant N-terminal tail of canonical H3) and in particular, lack of a conserved glutamine and a longer loop 1 region in the HFD (Malik & Henikoff, 2003). Typically, the use of these criteria has in the past allowed identification of even the more considerably divergent CenH3 proteins in some organisms (Malik & Henikoff, 2003). By applying these criteria to available genome assemblies or in the absence, by using transcriptomic screens to search for CenH3 transcripts, previous work led to the identification of putative CenH3 homologs in all monocentric insect orders that were analyzed (Diptera (flies and mosquitoes); Hymenoptera (wasps, bees, and ants); Coleoptera (beetles); Blattodea (cockroaches); Orthoptera (crickets); Phasmatodea (stick insects); and Ephemeroptera (mayflies). In strike contrast, a similar approach using all available genome sequences or in the absence, transcriptomes that collectively covered five insect orders comprising all four known holocentric insect lineages, CenH3 (and CENP-C) homologs could not be detected (Drinnenberg et al., 2014). Importantly, the branching order in the insect family tree as deciphered by comparing the phylogeny of the HFD that is conserved between CenH3 and other H3 proteins allowed confirmation that the absence of CenH3 in holocentric insects was due to recurrent *loss* in at least four independent lineages rather than the recurrent invention of or the horizontal transfer of CenH3 genes between monocentric clades (Drinnenberg et al., 2014). Furthermore, given the knowledge that holocentric insects are derived from monocentric ancestors, loss of CenH3 in every holocentric insect lineage suggested that dramatic changes in centromere architecture may have rendered this protein dispensable in the latter (Drinnenberg et al., 2014).

CenH3 is indispensable at the centromeres of most eukaryotes as evident in the catastrophic defects in chromosome segregation that follow upon its deletion in

several model organisms (M. D. Blower & Karpen, 2001; Buchwitz et al., 1999; Howman et al., 2000; Stoler et al., 1995; Talbert et al., 2002). Therefore, the centromere-specific function of CenH3 must be explored in order to perceive *why* the natural loss of this protein in some complex eukaryotes was indeed intriguing.

Overview of holocentricity in Insects

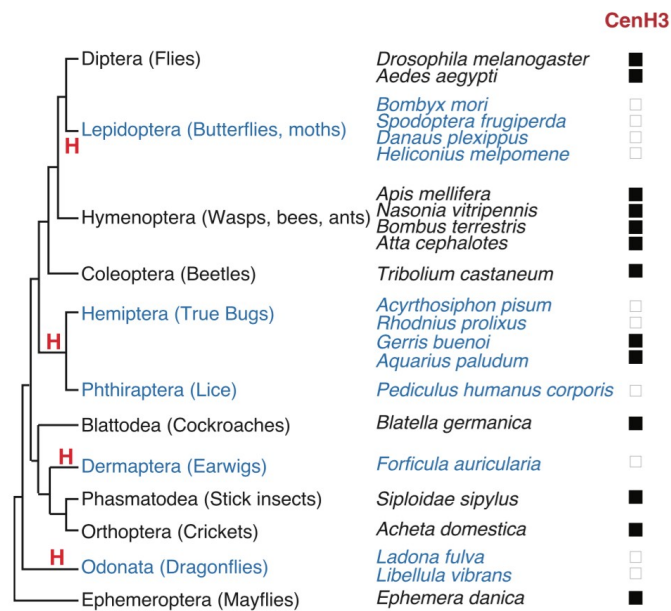
Holocentricity is a relatively common trait in insects. Based on phylogenetic studies, holocentricity in insects is believed to have evolved at least four independent times from monocentric ancestors (Drinnenberg et al., 2014). These four independent lineages comprise (i) the common ancestor of true bugs and lice (Hemiptera and Phthiraptera); (ii) butterflies and moths (Lepidoptera); (iii) dragonflies (Odonata); and (iv) earwigs (Dermaptera) (Figure 9).

The Hemiptera are by far the oldest group in which this trait was identified owing to the Schraders who greatly articulated this chromosome morphology in various insect specimens of the sub-order Heteroptera, which included broad-headed bugs and stink bugs, as well as in the sub-order Sternorrhyncha (coccids) (Hughes-Schrader & Schrader, 1961; Hughes-Schrader, Sally, 1944; Schrader, Franz, 1935). Thereafter, cytogenetic evidence (based on meiotic behaviors, X-ray fragmentation tests, EM studies and C (constitutive heterochromatin or centromere)-banding techniques) supporting holocentricity of many additional hemipteran insects is available, including in *Rhodnius proxilus* ((Buck, 1967)- cited from (Drinnenberg et al., 2014))), the principal vector of the parasite transmitting Chagas disease. Holocentricity is therefore considered to be a general characteristic of the Hemipteran insects (reviewed in (Drinnenberg et al., 2014)).

A wealth of similar cytogenetic studies describe holocentric chromosomes in multiple species of the Lepidoptera. Characterizing the chromosome morphology in species of this order is complicated by the characteristically small and numerous nature of chromosomes, where often chromosomes are visualized as tiny dots or rod-shaped structures under a microscope. Nevertheless, convincing cytogenetic evidence for holocentricity based on mitotic behavior and X-ray irradiation tests (Murakami & Imai, 1974) and more recently, IF-microscopy (Senaratne et al., 2020) is available for the domestic silkmoth, *B. mori* (Family: Bombycidae). Holocentricity is also convincingly reported for additional moth species from diverse families including the powdered quaker moth (Noctuidae), rusty tussock moth (Erebidae) and the Ailanthus silkmoth (Saturniidae), to name a few. There is also moderate evidence in some butterflies like the cabbage butterfly (Pieridae) (reviewed in (Drinnenberg et al., 2014)).

Overall, owing to the considerable sampling sizes within these two insect orders, a strong consensus has emerged that Hemiptera and Lepidoptera are representatively holocentric. In contrast, a few select species belonging to the remaining insect orders – namely, in the sister order to the Hemiptera- Phthiraptera; Odonata; and Dermaptera- have had their mitotic chromosome morphologies investigated, however is mostly restricted to cytological

observations (such as the lack of a primary constriction and the parallel migration of sister chromatids). Nevertheless, more convincing tests such as the mitotic behavior of irradiated chromosomal fragments (for the Phthiraptera); observation of obvious meiotic structures (for the Odonata); or the easily discernible primary constriction in the relatively fewer and longer chromosomes of some earwigs (Dermaptera) has allowed scientists to conclude holocentricity in some species belonging to these three orders. Additional sampling is however needed before concluding whether these insect orders are representatively holocentric (reviewed in (Drinnenberg et al., 2014)).



(Cortes-Silva et al., 2020)

Figure i9: Phylogeny of insects. Holocentric insect orders and species are indicated in blue. Inferred multiple transitions to holocentric chromosomes are labeled with “H.” Using protein homology searches of genomes or assembled transcriptomes, the ability and inability to find CenH3 is indicated by a black or white box, respectively.

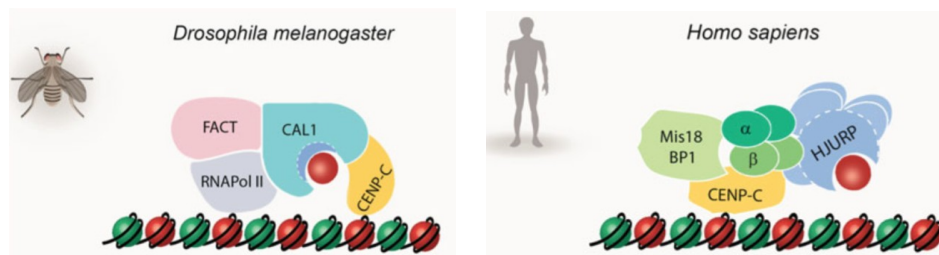
The role of CenH3 at centromeres

In most eukaryotes, CenH3 is specifically enriched in centromeric nucleosomes in place of canonical histone H3 that is incorporated into nucleosomes forming the bulk of the chromatin. In addition to demarcating the physical location of the centromere on chromosomes by the presence of CenH3, the memory of this location must be passed on to newly divided chromosomes. The ability to inherit the memory of the centromere location is essential, because with each round of DNA replication that happens once during each cell cycle, centromeric nucleosomes get equally dispersed into two daughter chromatids. Accompanying this equal partition is the reduction by half of CenH3-containing nucleosomes in each daughter chromatid and the gaps resulting from CenH3 dilution are occupied by histone H3.1 and H3.3 nucleosomes (Bodor et al., 2013; Dunleavy et al., 2011; Jansen et al., 2007; Ross et al., 2016). Continued dilution of CenH3 nucleosomes with the completion of each round of DNA replication without a means of replenishment would eventually lead a chromosome to lose its centromere identity. Therefore, it is essential that the centromere identity is stably propagated from one cell cycle to the next. The role of a centromere marker if not achieved by a genetic mechanism (such as the centromere-specific DNA sequence of budding yeast), which is inherited by daughter DNA strands upon DNA replication, must be achieved by other epigenetic (DNA sequence-independent) means (Karpen & Allshire, 1997).

Multiple lines of experimental evidence, for example in flies (Heun et al., 2006; Mellone et al., 2011; Mendiburo et al., 2011; Palladino et al., 2020; Roure et al., 2019) and in humans (Barnhart et al., 2011; Black et al., 2004; Carroll et al., 2010; Fachinetti et al., 2013; Guse et al., 2011; Jansen et al., 2007; H. Kato et al., 2013; Logsdon et al., 2015; Tachiwana et al., 2015) points to CenH3 being at the heart of centromere establishment, kinetochore formation, and stable propagation of centromere identity, thus making it the most upstream candidate for epigenetically marking the centromere in most eukaryotes.

CenH3 participates in a closed epigenetic loop that links newly-synthesized pre-nucleosomal CenH3 to “old” CenH3 present within centromeric nucleosomes at replicated centromeres from the previous cell cycle. Diverse organisms that have CenH3-defined centromeres use CenH3-specific chaperones to partially maintain the link between new and old CenH3 in order to propagate centromere identity across cell cycles by assembling new CenH3-containing nucleosomes. The CenH3-specific chaperone HJURP (Holliday Junction Recognition Protein) in humans, Scm3 in yeast (Dunleavy et al., 2009; Foltz et al., 2009) or CAL1 (chromosome alignment defect 1) in flies (Chen et al., 2014) directly binds pre-nucleosomal CenH3 through its N-terminal Scm3 domain (Scm3 domain is conserved among eukaryotes except for

Drosophila melanogaster CAL1 which has an Scm3-like domain at its N-terminus (Phansalkar et al., 2012; Sanchez-Pulido et al., 2009)). Upon binding, pre-nucleosomal CenH3 is sequestered for incorporation at the replicated centromere, which the CenH3-specific chaperone simultaneously recognizes by interacting directly with a network of centromere-associated components. A simplified version of this network consists only of the inner kinetochore protein CENP-C in flies (Heeger, 2005; Roure et al., 2019)(Figure i10, Left). CENP-C is the direct binding partner of CenH3 nucleosomes at the centromere (Carroll et al., 2010; H. Kato et al., 2013) and therefore can be considered as the “reader” of old, pre-existing CenH3 at centromeres. Whereas in humans, the network extends to (although is not limited to) both CENP-C and the Mis18 complex (containing Mis18-alpha, Mis18-beta and MIS18BP1/KNL2 (Mis18 binding protein1/kinetochore null 2)). The Mis18 complex mediates the link between pre-existing CenH3-bound- CENP-C and new CenH3-bound-HJURP (Figure i10, Right). Here, Mis18BP1 and Mis18-beta directly bind CENP-C (Dambacher et al., 2012; Moree et al., 2011; Stellfox et al., 2016); while the Mis18 complex also directly interacts with HJURP via HJURP’s centromere targeting domain within HCTD1 (HJURP carboxy terminal domain 1) (Barnhart et al., 2011; Fujita et al., 2007; Nardi et al., 2016; Wang et al., 2014) to mediate HJURP and CenH3 recruitment to centromeres. With the exception of budding yeast and *Drosophila*, the role of Mis18 proteins in the CenH3 deposition pathway is evolutionarily conserved from fission yeast to humans (Fujita et al., 2007; Hayashi et al., 2014; Maddox et al., 2007). In addition to the unique strategy for centromere inheritance in *Drosophila* discussed above that partly relies on CAL1 binding to CENP-C, the genetically inherited centromeres of budding yeast similarly rely on the coupling of CenH3-specific chaperones to the existing centromere to preserve CenH3 at the centromere locus (reviewed in (Zasadzinska & Foltz, 2017)).



(Zasadzinska & Foltz, 2017)

Figure i10: Protein complexes involved in CenH3 deposition pathway in *Drosophila* (left) and human (right). In *Drosophila*, CenH3 deposition requires active transcription mediated by the FACT (facilitates chromatin transcription) complex and RNA Pol II (Chen et al., 2014). CenH3 nucleosomes are depicted in red; canonical H3 nucleosomes are depicted in green.

More recently, CenH3 has also been implicated in a genetic basis of centromere identity. HJURP was originally named based on its in vitro ability to bind cruciform

DNA structures called Holliday junctions (Kato et al., 2007). Non-B form DNA conformations (such as DNA cruciforms) formed by centromere-specific sequences that are organized as dyad symmetries were recently reported to be enriched at the centromeres of many eukaryotes (Kasinathan & Henikoff, 2018). It has thus been speculated that non-B form DNA structures at eukaryotic centromeres may be recognized by HJURP and allow it to simultaneously deposit CenH3 at centromeres through its CenH3-specific chaperoning activity (Kasinathan & Henikoff, 2018).

Collectively, given the central involvement of CenH3 in maintaining centromere identity in diverse eukaryotes, CenH3 loss must be compensated for by an alternative strategy for the very same purpose. Together with its role in forming the foundation of kinetochore assembly in order to link DNA to microtubules, the identified loss of CenH3 in most holocentric insects was therefore intriguing in that it raised the most intuitive questions about the centromere biology of these organisms. For example,

(i) How are microtubule attachment sites delimited along the holocentric chromosomes of CenH3-deficient insects?

(ii) Is the memory of centromere location passed on to daughter cells and if so, how?

And further,

(iii) Given that CenH3 forms the foundation of the canonical kinetochore complex, has its loss also accompanied drastic changes to the kinetochores of holocentric insects?

I summarize below the studies in our lab focusing on this very last aspect (iii).

What we know about the kinetochore of CenH3-deficient holocentric insects

Initial studies on the kinetochore⁸ of CenH3-deficient holocentric insects were based on computational methods to search genome sequences and transcriptomes for conserved kinetochore components. Using this approach, several conserved components of the CCAN (inner kinetochore) and KMN network (outer kinetochore) could be identified (Drinnenberg et al., 2014). Conservation of inner and outer kinetochore components in CenH3-deficient holocentric insects suggested critical roles for identified proteins in the Lepidopteran chromosome segregation process and additionally to have kinetochores that are organized into a distinct inner and outer subunit that is composed of the same building blocks found in the kinetochores of CenH3-encoding organisms.

A recent study from our lab on the kinetochores of CenH3-deficient holocentric insects were conducted using cell lines derived from *B. mori* and *Spodoptera frugiperda* (Cortes-Silva et al., 2020), which belong to one of the four insect lineages (the Lepidoptera) in which CenH3 loss was identified (Drinnenberg et al., 2014). This recent study made use of stable cell-lines expressing tagged versions of the previously computationally-identified “bait” proteins (CENP-M, CENP-N, CENP-I of the inner kinetochore and Dsn1 and Nnf1 of the outer kinetochore) that were then used to conduct IP (immunoprecipitation) experiments to further profile their interaction partners by mass spectrometry. Mass-spectrometry analyses of pulled-down components revealed a generous CCAN in Lepidoptera which confirmed all previously predicted components. Importantly, mass-spec analyses combined with sensitive homology searches, also revealed an additional, previously un-identified protein that was pulled down with all bait proteins used- this was identified as the CENP-T protein of the CCAN (discussed below) (Cortes-Silva et al., 2020).

Extending on previous functional analyses (Mon et al., 2017), RNAi screens indeed verified that the identified components (particularly CENP-T and CENP-I of the inner

⁸ The kinetochore is a multi-protein complex that assembles specifically at centromeres (Fukagawa & Earnshaw, 2014). The kinetochore is composed of a distinct inner and outer subunit. The inner kinetochore subunit comprises a protein network known as the CCAN (constitutive centromere associated network) that is made of up to 16 different proteins present throughout the cell cycle. Several CCAN components can directly establish contacts to centromeres by physically interacting with CenH3 nucleosomes (Carroll et al., 2009, 2010; Falk et al., 2015; H. Kato et al., 2013) or by binding to DNA (Hori et al., 2008; Nishino et al., 2012; Sugimoto et al., 1994) to form the centromere-kinetochore interface. Upon cell division, the CCAN recruits the outer kinetochore subunit composed of the KMN network (Knl1, Mis12 and Ndc80 complex) (Musacchio & Desai, 2017), which interacts with microtubules to drive chromosome segregation (Monda & Cheeseman, 2018).

kinetochore and all outer kinetochore proteins) are essential for chromosome segregation in *B. mori*. Localization of CENP-T is dependent on several other inner kinetochore components (Cortes-Silva et al., 2020) and preliminary observations indicate an interdependent relationship between many CCAN factors (unpublished data from the lab). Further functional analyses focusing on identified CCAN components that are known to be more upstream in the kinetochore assembly pathway, specifically, the newly-identified Lepidopteran CENP-T protein, showed that CENP-T is not only essential for accurate mitosis, but is sufficient to direct complete outer kinetochore assembly in these organisms (Cortes-Silva et al., 2020).

The picture therefore emerges that in CenH3-deficient holocentric insects, in the event of losing CenH3-mediated kinetochore assembly as a means for physically connecting chromosomes to the spindle apparatus, this is compensated for by employing a CCAN-mediated pathway- potentially via its DNA-binding components such as CENP-T- to recruit the outer kinetochore and alternatively establish the required link to the spindle. The DNA-binding potential of Lepidopteran CENP-T was consistent with the identification of several positively charged amino acids in the conserved HFD at its C-terminus (Cortes-Silva et al., 2020). However, the DNA-binding potential of Lepidopteran CENP-T remains to be verified experimentally. Phylogenetic analyses additionally revealed that CENP-T is in fact retained in all other CenH3-deficient insect lineages (Cortes-Silva et al., 2020), indicating conserved critical functions for it such as in the aforementioned alternative pathway to the spindle and furthermore indicating that additional CENP-T-independent paths to the spindle are likely to exist in Lepidoptera, possibly via CENP-I (Cortes-Silva et al., 2020). Future studies will allow to dissect the hierarchy in this protein network even further.

Research aims

During my PhD research, I aimed to find answers to the former two of the basic questions listed previously, which interrogate the molecular architecture of the CenH3-deficient holocentromere:

(i) How are kinetochore assembly sites delimited along the holocentric chromosomes of CenH3-deficient insects?

(ii) Is the memory of centromere location passed on to daughter cells and if so, how?

I used cell lines derived from *B. mori* to carry out my experiments pertaining to these questions. My investigations were greatly facilitated by molecular tools that were developed in our lab, which included specific antibodies to some of the above-discussed Lepidopteran kinetochore components that were previously identified- in particular to the potentially DNA-binding kinetochore protein CENP-T.

Specifically, my research goals can be divided into four parts:

1) To map the centromere locations along the chromosomes of CenH3-deficient Lepidoptera **(Result 1)**

2) To characterize the DNA sequence components underlying the identified centromere sites **(Result 2)**

3) To characterize the chromatin components underlying centromere sites **(Result 3)**

4) To carry out functional analyses on the necessity of identified components for centromere specification in Lepidoptera **(Results 4(i) & 4(ii))**

Importantly, as opposed to previous reports of isolated cases of CenH3 loss in primitive eukaryotic lineages like the kinetoplastids (Akiyoshi & Gull, 2014; Talbert et al., 2009) and some early-diverging fungi (Hooff et al., 2017; Navarro-Mendoza et al., 2019), the concomitant occurrence of CenH3 loss and the presence of a holocentric architecture in at least four independently-derived insect lineages- each comprising multiple species with confirmed cases of CenH3 loss- is the first time a general correlation could be made between the loss of this protein and a change in centromere organization spanning an evolutionary timescale. This therefore also highlights the need to revisit the current evolutionary standpoint of the origins of holocentricity in itself. I address this point within the context of my results in the discussion.

Results

Result 1: The kinetochore forms a broad localization pattern along *B. mori* mitotic chromosomes

To understand holocentromere architecture in CenH3-deficient insects, we used a cell line derived from our representative insect model system, *B. mori*. We targeted one previously identified centromere-proximal component of the *B. mori* kinetochore, CENP-T (Cortes-Silva et al., 2020) to visualize its localization pattern on mitotic *B. mori* chromosome spreads by IF microscopy using a custom-made antibody (Cortes-Silva et al., 2020). We observed that the CENP-T-specific IF signal formed broad localization patterns along the polar length of sister chromatids (Figure 1A), a pattern reminiscent to kinetochore staining in other holocentric organisms (Buchwitz et al., 1999). While CENP-T is an inner kinetochore component that binds directly to DNA as shown in vertebrates (Hori et al., 2008), the centromere-distal outer kinetochore components bridge centromeric DNA to the spindle apparatus, thus serving as a proxy for potential sites of spindle fiber attachment (Musacchio & Desai, 2017). To test whether the broad localization pattern we observed for CENP-T corresponds also to sites of outer kinetochore assembly, we co-stained our chromosome spreads with another custom-made antibody specific for the outer kinetochore component Dsn1 (Cortes-Silva et al., 2020). We found that the CENP-T and Dsn1 immunosignals co-localize with one another (Figure 1B), demonstrating that diffuse regions along the poleward surface of *B. mori* sister chromatids represent potential sites driving chromosome segregation during mitosis. These stainings are the first visualizations of kinetochore components along *B. mori* chromosomes, thereby building on previous cytological data (Murakami & Imai, 1974) and confirming that *B. mori* chromosomes are holocentric.

Result 2: Half of the *B. mori* genome is permissive for kinetochore assembly

We proceeded to perform X (cross-linked)-ChIP-seq experiments targeting CENP-T in unsynchronized *B. mori* cell populations in order to obtain genomic resolution maps of CENP-T's distribution on chromatin. These analyses revealed broad regions of CENP-T enrichment (Figure 1C) with maximum enrichment scores of about 2-fold (log₂) over input, indicating an extensive, yet low level of CENP-T localization along chromosomes. This pattern was highly reproducible across replicates (Figure S1A and S1B). Given that the ChIP-seq libraries were prepared from asynchronous cell populations containing ~ 2% mitotic cells (Cortes-Silva et al., 2020) and that the *B. mori* CENP-T, as the vertebrate CENP-T homologs (Hori et al., 2008) also localizes to

chromatin during interphase (Figure S1E), CENP-T interphasic localization likely contributes to the majority of its enrichment signal.

To evaluate the extent to which CENP-T-enriched sites identified by ChIP-seq are functional kinetochore assembly sites to drive chromosome segregation during mitosis, we performed X-ChIP-seq to map the genome-wide distribution pattern of Dsn1. As other outer kinetochore components, Dsn1 is found on centromeric chromatin only during mitosis (Musacchio & Desai, 2017). Analogous to our co-stainings of CENP-T and Dsn1 on mitotic chromosome spreads, the ChIP-seq profile of Dsn1 also correlated well with that of CENP-T ($r = 0.8$) (Figure 1C, G).

To better characterize CENP-T-binding patterns, we used a custom domain-calling pipeline to annotate CENP-T domains. This method revealed that CENP-T domains show no obvious clustering towards the center or telomeric regions of chromosomes (Figure 1D); are of variable size (median size = 36 kb); and have a genome-wide median coverage of 54% (Figure 1E). Considering that CENP-T coverage is quantified for CENP-T ChIP-seq signal from a cell population, this number reflects the average CENP-T coverage for that population. Thus, while approximately half of the genome is CENP-T-permissive, it is possible that from one cell to the next, CENP-T occupies only a fraction of permissive sites.

Next, we analyzed the distribution of repeat sequences underlying CENP-T sites. Approximately 47% of the *B. mori* genome is comprised of repetitive DNA sequences, with around 97% of these repeats being Class I and Class II TEs (Kawamoto et al., 2019). To determine if CENP-T domains preferentially occur in these repetitive elements, we first intersected our annotated CENP-T domain coordinates with a database of consensus transposon sequences for *B. mori* (kind gift from the Pillai Lab, University of Geneva). This approach revealed that the proportion of transposon sequences underlying CENP-T domains is not significantly different to that in the rest of the genome (Figure 1F). Second, as a complementary approach, we searched de novo for any centromere-enriched repeats as described (Smith et al., 2020). Consistent with the first approach, these analyses also revealed that CENP-T-permissive sites in the *B. mori* genome are not enriched for repetitive DNA sequences (Figure S1F). Additionally, intersecting the CENP-T domain coordinates with *B. mori* gene annotations (obtained from SilkBase: <http://silkbases.ab.a.u-tokyo.ac.jp>) revealed that CENP-T domains are relatively depleted in genes (Figure 1F). Collectively, these results led us to conclude that *B. mori* centromeres are organized as broad domains that are composed of complex DNA and depleted from gene bodies.

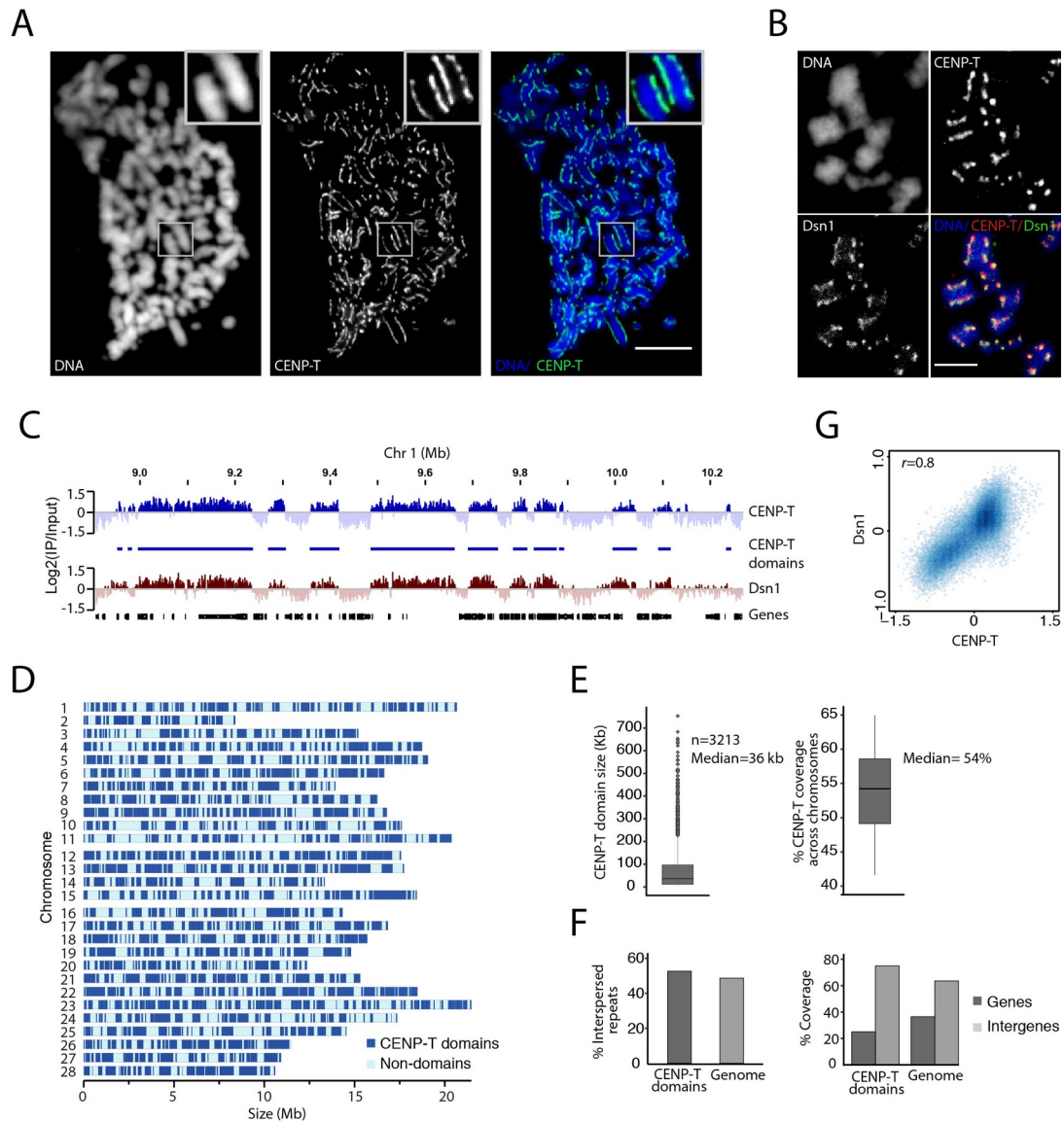


Figure 1: Kinetochore localization patterns in *B. mori*.

A) Representative IF image of *B. mori* DAPI (4',6-diamidino-2-phenylindole)-stained mitotic chromosomes (blue) showing broadly-distributed immunosignal patterns of CENP-T (green). Scale bar: 5 μ m. **B)** Representative IF image of *B. mori* DAPI-stained mitotic chromosomes (blue) showing immunosignal patterns of CENP-T (red) and Dsn1 (green). Co-localization of CENP-T and Dsn1 signals can be distinguished as yellow foci in the overlay. Scale bar: 5 μ m. **C)** Genome-browser snapshot of a representative portion of *B. mori* chromosome 1 for CENP-T X-ChIP-seq, CENP-T domains, Dsn1 X-ChIP-seq and annotated genes. ChIP-seq signals are represented as histograms of the average log₂ ratio of IP/Input in genome-wide 1 kb windows. **D)** Size-scaled schematics of 28 *B. mori* chromosomes showing the distribution of CENP-T domains (dark blue segments). **E)** Features of CENP-T domains: boxplots showing the sizes of CENP-T domains (left) and the genome coverage in percent by CENP-T domains (right). **F)** Genomic features underlying CENP-T domains: barplots showing the fraction in percent of annotated interspersed repeats within CENP-T domains (dark grey) and

genome-wide (light grey) (left); and the fraction in percent of annotated genes within CENP-T domains (dark grey) and genome-wide (light grey) (right). **G**) Genome-wide correlation plot of CENP-T and Dsn1 occupancy. Average log₂ ratios of IP/Input in 10 kb windows were used for plotting and calculating the Pearson correlation coefficient (r).

Result 3: Kinetochores attachment sites in *B. mori* are anti-correlated with actively transcribed chromatin

Given the broad distribution pattern and absence of any consensus sequence underlying CENP-T sites, it is unlikely that centromeres along *B. mori* chromosomes are defined by a specific DNA sequence. Our use of genomics approaches to profile the underlying chromatin environment instead revealed several correlations in the distribution of CENP-T with respect to chromatin marks governing transcriptional status. We found a positive correlation between the distributions of CENP-T and H3K27me₃ (tri-methylated Lysine 27 on histone H3) ($r = 0.6$) (Figure 2A, B), a histone mark that is typically associated with transcriptionally-silent heterochromatin (Kouzarides, 2007). Conversely, we found a negative correlation in the distributions of CENP-T and two histone marks that are associated with a transcriptionally-active chromatin state (Kouzarides, 2007): (i) H3K4me₃ (tri-methylated Lysine 4 on histone H3) ($r = -0.2$), and (ii) H3K36me₃ (tri-methylated Lysine 36 on histone H3) ($r = -0.6$) (Figure 2A, B). We also profiled the distribution of H3K9me₂ and H3K9me₃ (di- and tri-methylated Lysine 9 on histone H3, respectively), two histone marks that are typically associated with transcriptionally silent heterochromatin (Kouzarides, 2007). However, the resulting ChIP-seq profiles generated using multiple different antibodies and ChIP protocols for H3K9me_{2/3} were very similar to a histone H3 ChIP-seq profile ($r = 0.8$ for H3K9me₂ vs H3; and $r = 0.7$ for H3K9me₃ vs H3, respectively) (Figure S2B, C). This offered us limited confidence in the observed patterns of H3K9me_{2/3} in our cell line. However, as opposed to the H3K9me₃-enriched heterochromatin blocks found at the pericentromeres of monocentric organisms (Sullivan & Karpen, 2004), a lack of such regions was observed in our *B. mori* cell line as evident in the absence of any chromocenters in interphase nuclei (Figure S2D).

We additionally generated mRNA-seq (messenger RNA sequencing) data for our cell line in order to assess the expression levels at *B. mori* kinetochores attachment sites. We found that mapped transcripts negatively correlated with CENP-T domains (Figure 2A). Additionally, annotated genes that did fall completely within CENP-T domains had significantly lower expression levels (median = 0.06 normalized expression units) compared to the genome-wide average (median = 1.53 normalized expression units) (Figure 2C). To account for transcripts mapping to non-annotated genes, we also looked at the expression levels across the entire genome. As with the gene-level analysis, we found that genomic regions that are enriched for CENP-T

ChIP-seq signal have significantly lower expression values (median = 0.2 normalized expression units) than CENP-T- depleted regions (median = 3.06 normalized expression units) (Figure 2C). Consistent with our ChIP-seq profiles for histone marks and our mRNA-seq profiles, the ChIP-seq profile of RNA Pol II also revealed a negative correlation with CENP-T throughout the genome ($r = -0.64$) (Figure 2A, B). Finally, we also profiled by ChIP-seq epitope-tagged H3.3 (histone 3.3) driven under a constitutive promoter in order to map regions of nucleosome turnover (Kraushaar et al., 2013). Here, we found that H3.3-enriched sites show a stronger anti-correlation with CENP-T ($r = -0.76$) than the active histone marks and RNA Pol II profiles (Figure 2A, B). Taken together, our results indicate that *B. mori* centromeres are excluded from genomic regions undergoing active nucleosome turnover, driven by transcription or other chromatin remodeling processes (Figure 2A, B).

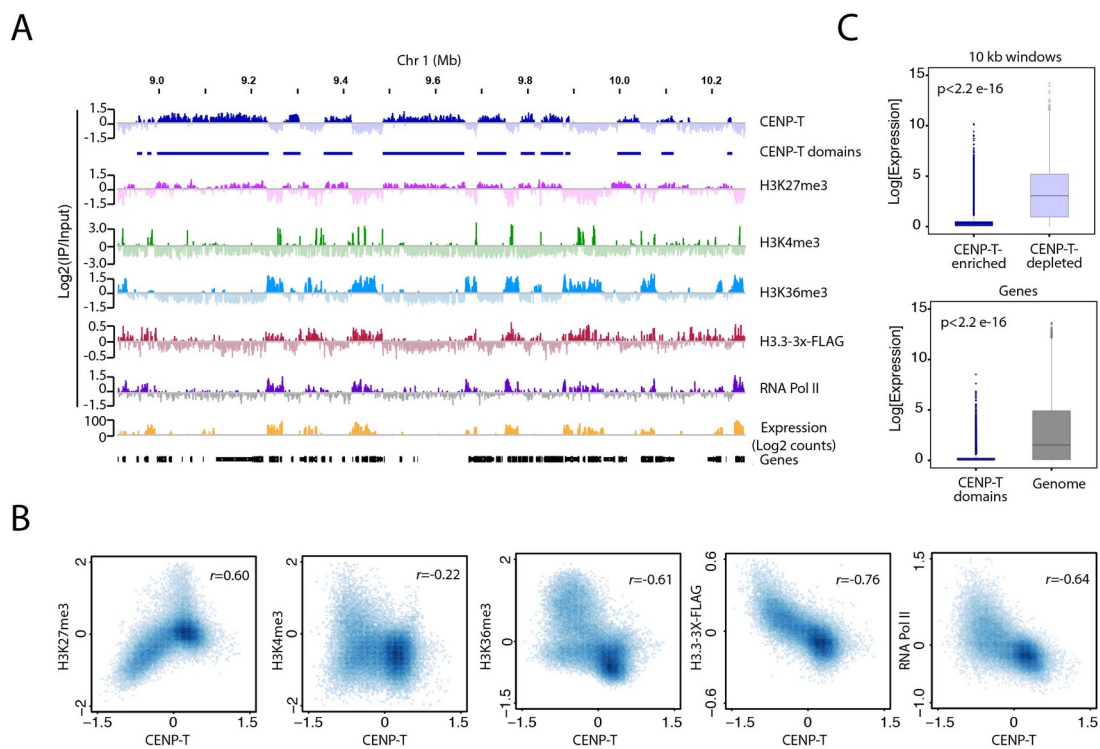


Figure 2: *B. mori* kinetochore attachment sites are anti-correlated with actively transcribed chromatin and nucleosome turnover.

A) Genome-browser snapshot of a representative portion of *B. mori* chromosome 1 for CENP-T X-ChIP-seq, CENP-T domains, H3K27me3 N (native)-ChIP-seq, H3K4me3 X-ChIP-seq, H3K36me3 N-ChIP-seq, H3.3-3X-FLAG N-ChIP-seq, RNA Pol II X-ChIP-seq, mRNA-seq and annotated genes. ChIP-seq signals are represented as histograms of the average log₂ ratio of IP/Input in genome-wide 1 kb windows. mRNA-seq signal is represented as a histogram of log₂ transformed BPM (bins per million mapped reads). **B)** Genome-wide correlation plots comparing the occupancy of CENP-T and H3K27me3,

H3K4me3, H3K36me3, H3.3-3X-FLAG and RNA Pol II. Average log₂ ratios of IP/Input in 10 kb windows were used for plotting and calculating the Pearson correlation coefficient (r). Comparisons between N-ChIP-seq and X-ChIP-seq or X-ChIP-seq replicates for the histone marks are shown in Figure S2A. C) Top: boxplot showing the expression levels in genome-wide 10 kb windows that are CENP-T-enriched (blue) or CENP-T-depleted (purple), and bottom: boxplot showing the expression levels in annotated genes that are 100% within CENP-T domains (blue) as compared to annotated genes genome-wide (grey). Expression levels are represented as the log₂ transformed TPM (transcripts per million mapped reads). Statistical significance was tested using the Kolmogorov-Smirnov test.

Result 4 (i): Hormone-induced perturbations of gene expression result in proximal CENP-T loss or gain

To test our hypothesis that there is a causal link between *B. mori* centromere location and chromatin activity, we used an approach that allowed us to study the effects of induced transcriptional changes on CENP-T localization patterns. In insects, the ecdysone hormone response is a well-defined transcriptional response leading to the systematic activation or repression of a subset of specific genes involved in metamorphoses and development (Yamanaka et al., 2013).

We treated asynchronous *B. mori* cells with 20E (20-Hydroxyecdysone) for forty-eight hours and carried out mRNA-seq to determine whether our cell line had an effective transcriptional response (Figure 3A). Indeed, we found that while the expression pattern in annotated genes remained overall well-correlated before and after 20E treatment ($r = 0.99$), several genes showed differential expression upon treatment (as indicated by clusters forming away from the diagonal in the +/- 20E correlation scatterplot; Figure 3B). To identify a subset of differentially expressed genes for further analyses, we applied cut-offs in expression level to define up- and down-regulated genes (See Methods). Among a subset of twenty-six up-regulated genes, we found known 20E-responsive genes including orphan nuclear receptor genes (Yamanaka et al., 2013)(Figure 3B). We also identified one down-regulated gene that corresponded to a 20E-hydroxylase (Figure 3B). The functional annotations of these genes furthered our confidence in the specificity of the 20E response in our cell line.

Having confirmation of a visible change in expression, we next profiled CENP-T occupancy under the same conditions (Figure 3A). To identify any changes in CENP-T localization, we compared CENP-T enrichment patterns across genome-wide 10 kb windows between treated and untreated conditions. Consistent with the largely unaltered transcriptome in annotated genes (Figure 3B) and 10 kb windows (Figure 3D), global CENP-T occupancy levels were also well-correlated in treated and untreated conditions ($r = 0.9$) (Figure 3C). Nevertheless, we were able to identify

specific loci with altered CENP-T levels after treatment (Figure 3C). Comparing mRNA expression levels in genomic regions showing lost or gained CENP-T binding after 20E treatment revealed that CENP-T depleted regions showed elevated expression while CENP-T enriched regions showed decreased levels of expression (Figure 3D). This is in agreement with our above findings that CENP-T binding is more robust in transcriptionally silent regions.

Next, we zoomed in on each of the individual loci that lost or gained CENP-T after treatment in order to further interpret each CENP-T loss or gain event with respect to the broader chromosomal environment. We found that the most pronounced cases of CENP-T loss (representing seventeen out of forty-eight CENP-T-depleted 10 kb genomic windows) were in fact in the proximity of two up-regulated genes. These loci corresponded to large consecutive regions of CENP-T loss after 20E treatment on chromosomes 10 and 15 where in each case, the losses were just upstream of the genes encoding for a 20E-specific nuclear receptor protein (Figure 3E). On the other hand, when we zoomed in on a genomic region on chromosome 7 that gained CENP-T after 20E treatment, we observed the opposite scenario, where the gain in CENP-T was in close proximity to a down-regulated 20E-hydroxylase gene (Figure 3E). While we cannot explain all of the observed changes in CENP-T occupancy after 20E treatment, the above described examples of CENP-T gain or loss near genes with altered expression allowed us to partially link CENP-T localization patterns to changes in transcriptional activity. The fact that altered CENP-T occupancy did not necessarily co-localize with the altered transcriptional output, but rather extended to large upstream regions, reinforces our hypothesis that changes of the chromatin landscape such as nucleosome eviction or reassembly in gene bodies, promoter or enhancer regions of up- or down regulated genes interfere with CENP-T localization.

In order to further evaluate whether it is a change in chromatin dynamics in the proximity of differential-regulated genes that underlies the changes in CENP-T occupancy, we allowed the transcriptional program to reset over the course of five days upon removal of the 20E hormone from the media (Figure 3A). Analyses of gene expression levels in the hormone-washout sample revealed that the previously identified subset of induced genes was mostly restored to expression levels similar to that of the control gene subset (Figure S3A). To evaluate the effects of restored gene expression on CENP-T localization, we once again profiled CENP-T occupancy 10 days following hormone-washout (Figure 3A). We found that upon 20E-washout, the prominent CENP-T losses on chromosomes 10 and 15 were recovered to levels similar to wild-type profiles (Figure 3E). The recovered events of previous CENP-T loss extended upstream of the now repressed mRNA output, once again indicating a link between chromatin activity and CENP-T deposition. The CENP-T gain upon 20E-treatment that was proximal to the down-regulated 20E-hydroxylase gene on chromosome 7 did not completely recover, which could be explained by incomplete

restoration of expression of this gene (Figure 3E). In addition to these individual cases, similar levels of recovery following 20E-washout were observed genome-wide for regions with differential CENP-T occupancy (Figure S3B) along with comparable expression levels across all genomic windows (Figure S3C).

Taken together, our results support the hypothesis that the dynamics of the chromatin landscape determined by promoter activation or RNA Pol II passage govern CENP-T occupancy such that CENP-T is removed from active chromatin. The complete loss of CENP-T upon promoter and gene activation also shows that no immediate recycling mechanism exists to restore CENP-T levels in those regions. Finally, the recovery of CENP-T localization upon 20E-washout also indicates that restoring low chromatin dynamics is sufficient for centromere formation.

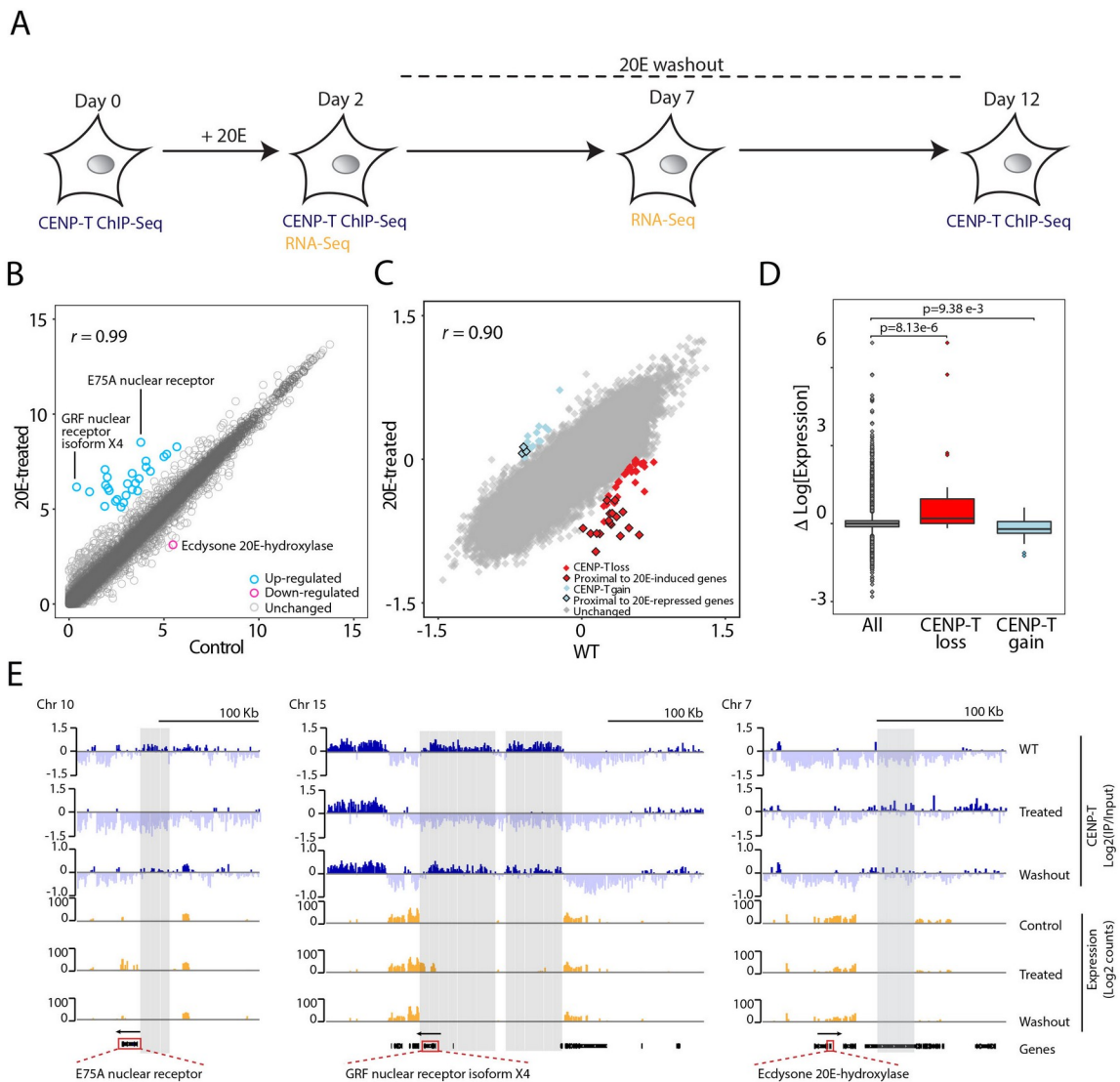


Figure 3: Hormone-induced perturbations of gene expression result in proximal CENP-T loss or gain.

A) Schematic summarizing the steps of 20E-treatment and 20E-washout. **B)** Genome-wide correlation plot comparing the expression levels of annotated genes in 20E-treated vs DMSO (dimethyl sulfoxide) control. Twenty-six differentially-expressed genes are highlighted. Functions of two up-regulated genes and one down-regulated gene that could be linked to three cases of differential CENP-T occupancy in the 20E-treated condition are annotated. Log₂ transformed TPM is used to represent expression level per gene and to calculate the pearson correlation coefficient (r). **C)** Genome-wide correlation plot comparing CENP-T occupancy before (WT (wild type)) and after 20E-treatment in 10 kb windows. 10 kb windows with differential CENP-T occupancy are demarcated. Average log₂ ratio of IP/Input in 10 kb windows is plotted and used for calculating the pearson correlation coefficient (r). Only those 10 kb windows with total loss or total gain of CENP-T after 20E treatment from a previously enriched or depleted state, respectively were considered. **D)** Boxplot: difference in expression levels between 20E treatment and control in genome-wide 10 kb windows (grey) and the subset of 10 kb windows with depleted or enriched CENP-T occupancy (red and blue, respectively). Difference in expression was calculated by subtracting the log₂ transformed TPM scores of control from 20E-treated for each 10 kb window. Statistical significance was tested using the Kolmogorov-Smirnov test. **E)** Genome-browser snapshots of *B. mori* chromosomes showing two cases of CENP-T loss (left & middle) and one case of CENP-T gain (right) that could be linked to proximal changes in gene expression. Tracks: CENP-T X-ChIP-seq profiles (blue) and RNA-seq (RNA-sequencing) profiles (orange) for WT (DMSO control for RNA-seq), 20E-treated and 20E-washout. Pre-identified 10 kb windows with differential CENP-T occupancy (grey rectangles) fall consecutively within these regions. The three differentially-expressed 20E-specific genes lying upstream of these regions are annotated. ChIP-seq signal is represented as the log₂ ratio of IP/Input in genome-wide 1 kb windows. RNA-seq signal is represented as log₂ transformed BPM. CENP-T occupancy around the remaining twenty-three pre-identified differentially-expressed genes are shown in Figure S3D.

Result 4 (ii): Differentially expressed orthologous genes in Lepidoptera show opposite patterns of CENP-T localization

Given that the same chromosomal locus could be made permissive or repressive to CENP-T by merely changing the underlying chromatin activity status, we reasoned that such opposite patterns of CENP-T enrichment should be readily observable in an experimentally unperturbed setting, for example, on a gene that naturally varies in expression levels. To address this possibility, we turned to a close Lepidopteran relative to *B. mori*, the cabbage looper *Trichoplusia ni*. We hypothesized that differentially-expressed orthologous genes between these two species should provide

a natural template of varying expression levels at orthologous loci to which we could correlate CENP-T occupancy levels (Figure 4A).

To first evaluate whether centromeres in *T. ni* are defined in a similar way to *B. mori*, we profiled CENP-T in an unsynchronized germ cell line derived from *T. ni* (Granados et al., 1986). We found that the distribution of *T. ni* CENP-T X-ChIP-seq signal resembled that of *B. mori*, wherein CENP-T localized to broad chromosomal regions (Figure S4A). Next, we did ChIP-seq profiling of the histone marks H3K27me3 and H3K36me3 in our *T. ni* cell line, which we used as markers of transcriptionally-repressed and -active chromatin, respectively. Similar to *B. mori*, in *T. ni*, the genome-wide distribution of CENP-T was positively and negatively correlated with H3K27me3 ($r = 0.5$) and H3K36me3 ($r = -0.5$), respectively (Figure S4A, B) and showed a negative correlation to mapped mRNA transcripts (Figure S4A). Thus, we concluded that the CENP-T localization in *T. ni* is correlated with silent chromatin- similar to what is seen in *B. mori*.

We then reciprocally searched the *B. mori* and *T. ni* proteomes with one another to select a set of encoding orthologous genes. This revealed > 9500 orthologs to which we applied a fold-change and cut-offs in expression to select a subset of differentially-expressed genes for each species (See Methods). Accordingly, we selected 115 genes that are expressed in *B. mori* (and low in *T. ni*) and 110 genes that are expressed in *T. ni* (and low in *B. mori*) (Figure 4B). We then quantified the average CENP-T ChIP-seq scores for *B. mori* and *T. ni* over both sets of genes. In line with our previous observations, we found that CENP-T ChIP-seq signal was depleted over those genes that are expressed in *B. mori*, while CENP-T ChIP-seq signal was enriched over the corresponding lowly expressed orthologs of *T. ni* (Figure 4C). In a similar manner, we observed that CENP-T ChIP-seq signal was enriched over those genes that are lowly expressed in *B. mori*, while CENP-T ChIP-seq signal was depleted over the corresponding expressed orthologs of *T. ni*. We have thus demonstrated that across orthologous genes, CENP-T occupancy is significantly different only in those cases where there is differential expression, further supporting our hypothesis linking centromere formation to the underlying chromatin activity status.

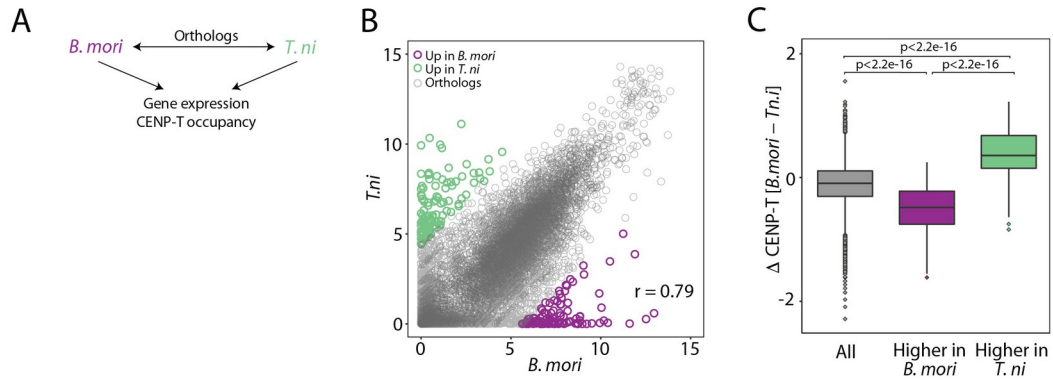


Figure 4: Differentially expressed orthologous genes in Lepidoptera show opposite patterns of CENP-T localization.

A) Schematic summarizing the concept to identify a link between CENP-T occupancy and transcriptional activity by using naturally differentially-expressed orthologous genes of *B. mori* and *T. ni* as a template. **B)** Correlation plot showing the expression level of orthologous genes in *B. mori* and *T. ni*. Genes with similar expression levels (grey circles) and those corresponding to the subsets of the most highly expressed genes that were identified in each species, which are silent in the corresponding species (purple or green circles) are indicated. Expression level per gene is represented as a log₂ transformed TPM score and was used to calculate the Pearson correlation coefficient (r). **C)** Boxplot: difference in CENP-T occupancy across all orthologs (grey) and across the subset of highly expressed genes of *B. mori* and *T. ni* (purple and green, respectively). Difference in CENP-T occupancy per gene was calculated by subtracting the average log₂ IP/Input scores for CENP-T ChIP-seq in *B. mori* from that of *T. ni*. Statistical significance was tested using the Kolmogorov-Smirnov test.

Discussion

The results of my PhD project unravelled the CenH3-deficient holocentromere architecture of Lepidoptera, which allowed us to make an important link between their chromatin landscape and centromere distribution. Our use of a perturbation system to show that this landscape is critical for creating centromere-permissive or -dismissive environments along Lepidopteran chromosomes provides insights into a new mode of centromere definition independent of CenH3.

More precisely, we identified recurrent patterns of negative correlations between the genome-wide distribution of CENP-T and factors associated with active chromatin. This included RNA Pol II, the passage of which through a DNA template is known to disrupt and mobilize nucleosomes in its wake during transcription (Teves & Henikoff, 2014); as well as the histone H3 variant H3.3, which is deposited in a replication-independent manner at regions undergoing replication-uncoupled nucleosome replacement (Ahmad & Henikoff, 2002)(such as during transcription) (Torné et al., 2019). Both of these are good indicators of chromatin disruption associated with transcriptional elongation activity. Importantly, our study allowed us to link CENP-T localization specifically to chromatin activity resulting from nucleosome turnover and not transcription for the following reasons: Although we could identify a negative correlation to other proxies of gene activity like (i) RNA-seq data; and (ii) to histone marks present on active gene bodies (H3K36me3), H3.3 reflected the strongest anti-correlation to CENP-T. H3.3 enrichment is not just a consequence of transcriptional elongation; rather, can also mark dynamic regions outside of active genes and can reflect histone replacement in additional dynamic regions throughout the genome (Deal et al., 2010; Deaton et al., 2016; Kraushaar et al., 2013). Additionally supported by the observation that induced CENP-T loss/gain occurred in regions upstream of annotated genes that are likely to comprise gene regulatory elements, a direct link to dynamics resulting from nucleosome turnover could be made with transcription or RNA Pol II passage being one of the contributing factors.

Thus, we propose that Lepidopteran kinetochores can assemble non-specifically and anywhere along the chromosomes where nucleosome turnover is low rather than being dependent on an active recruitment process involving a centromere-specific epigenetic or genetic component. Changes to the chromatin landscape, resulting for example from changes in gene expression can thereafter disrupt kinetochore attachment leading to its complete loss due to the absence of an active recycling mechanism. The CENP-T profile that we thus measure corresponds to attached kinetochores that could persist over the cell cycle. This is different to the dynamics of H3.3 during transcription for instance, where a combination between new H3.3 deposition and recycling of pre-existing H3.3 coordinated by its dedicated chaperone,

the HIRA (histone regulator A) complex enables the maintenance of the epigenetic information at that locus (Torné et al., 2019). In contrast, the presence of the lepidopteran kinetochore is not memorized and can change from cell cycle to cell cycle dependent on the chromosome-wide chromatin landscape (Figure 5).

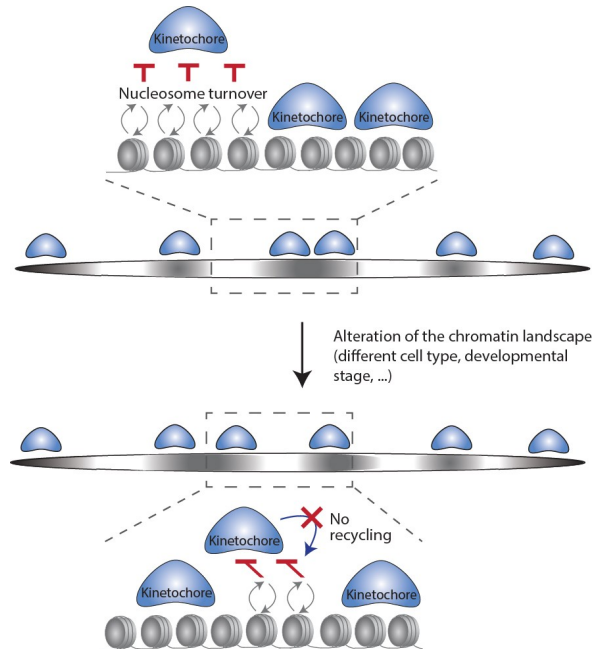


Figure 5: Model for the architecture of CenH3-deficient holocentromeres in insects. Kinetochores bind chromosome-wide due to lack of centromere specificity. Binding is opposed only by high nucleosome turnover. Any changes to the chromatin landscape that result in alterations to the nucleosome turnover profile will also alter kinetochore attachments toward more stable regions. This kinetochore binding profile is not epigenetically inherited from one cell cycle to the next, therefore resulting in a different kinetochore binding pattern for every different nucleosome turnover profile.

The apparent dependency on nucleosome turnover for centromere localization implies a challenge for holocentric organisms such as Lepidoptera and *C. elegans* in that their chromosome segregation efficiency is likely to directly depend on being able to establish a minimum number of stably attached kinetochores chromosome-wide. Thus, holocentromeres could be relatively more sensitive to any stimuli that can globally increase or decrease chromatin dynamics to the extent of negatively impacting chromosome segregation fidelity. Therefore, as opposed to inducing local changes to the transcriptional landscape, it would be interesting to test this hypothesis by applying a non-specific stress such as an environmental change (eg. a gradient of temperatures) that could globally affect the nucleosome turnover profile to various extents and to study its consequences on holocentric chromosome segregation fidelity. It would be interesting to compare these results to similar experiments in monocentric organisms which have a specific and therefore perhaps more robust centromere definition mechanism that might tolerate better the impacts of such global changes on the chromosome segregation efficiency. Given the absence of an active centromere specification mechanism, holocentric *B. mori* would therefore serve as an ideal model system in which to characterize the role of environment in centromere fitness for the first time.

Holocentromere regulation further emphasizes centromere plasticity

The organization and inheritance of the Lepidopteran holocentromere might be conceptually similar to the CenH3-encoding holocentromere in *C. elegans*. In this organism, it was proposed that in later embryonic stages and adult worms, CenH3 incorporation into transcriptionally-silent chromatin may be a result of low nucleosome turnover in those regions, which delimited CenH3 to these sites by way of exclusion from regions with high nucleosome turnover (Steiner & Henikoff, 2014). CenH3 nucleosomes are thus proposed to stably remain over chromosomal domains with low levels of RNA Pol II occupancy. Given the observation that *C. elegans* CenH3 was completely turned-over at each cell cycle and discontinued in the germline, the authors further proposed that CenH3 might not propagate centromere identity in this organism (Gassmann et al., 2012). The ubiquitous negative relationship between active chromatin and centromeres throughout the genomes of CenH3-deficient Lepidoptera closely resembles that of *C. elegans*. That is, based on our chromatin-perturbation experiments, we can conclude that chromatin activity (such as by way of transcription) is sufficient to induce complete dissociation of CENP-T (Figure 3). In particular, the 20E washout experiment further showed that CENP-T can thereafter accumulate de novo, with no pre-existing requirement of CENP-T as a cue. The potential similarities in centromere regulation regardless of the absence or presence of CenH3 highlight the relevance of chromatin dynamics for holocentromere organization across species. In the broader context, it also highlights the degree of centromere plasticity across eukaryotes.

Adding it all up...

The relevance of chromatin dynamics for shaping holocentromere architecture further stands out when we correlate the characteristics of annotated centromere domains in both *C. elegans* (Gassmann et al., 2012) and *B. mori* (this study) to the respective fraction of active compartments in each genome. The *C. elegans* genome size is 100 mb (*C. elegans* Sequencing Consortium 1998), ~4.6X smaller than that of *B. mori* (~460 mb) (Kawamoto et al., 2019). Given its relatively smaller size, it is also not surprising that protein-coding genes in *C. elegans* results in a higher gene density per genomic window, which we approximate to be ~6.6 X higher than that of *B. mori* based on published statistics (Table 1). This therefore reveals a significant level of gene compaction in *C. elegans*, as also evident in the fact that it has one of the smallest average intron sizes among eukaryotes (Suetsugu et al., 2013). *C. elegans* (and other nematodes) are in fact among the few eukaryotes to have genes organized in clusters (operons), which constitutes 15% of all its genes (Lercher, 2003). It is

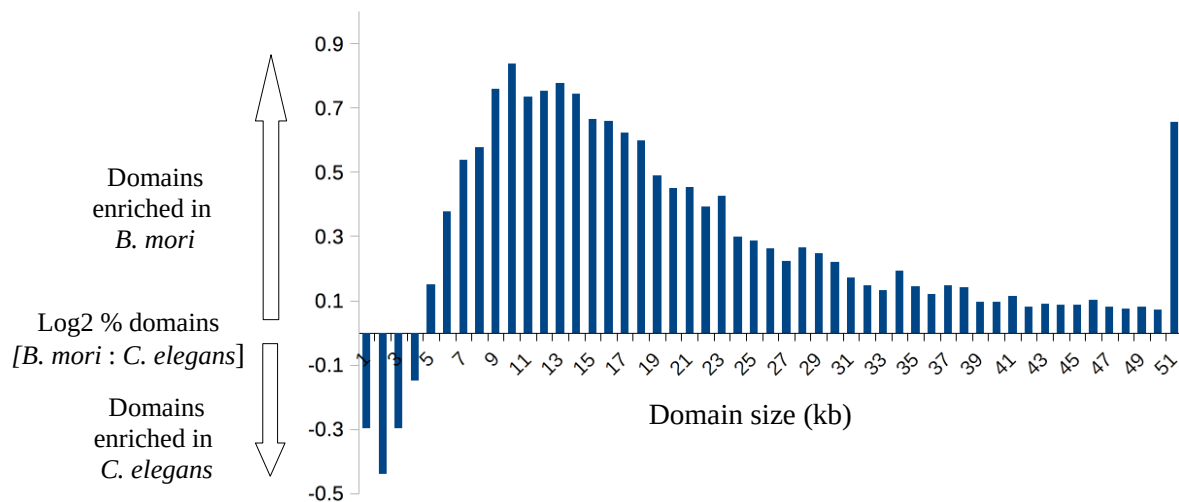
known that the gene distribution in *C. elegans* is biased, where the distal chromosomal arms have longer genes with larger introns while the chromosome centers have shorter genes that are more highly expressed (Ahringer & Gasser, 2018). This gene expression pattern is reflected in its chromatin and centromere landscape, where silent histone marks like H3K27me3, H3K9me2/3 as well as CenH3 domains are enriched in the distal arms while active histone marks like H3K36me3 are enriched in the center (Ahringer & Gasser, 2018; Ho et al., 2014; Steiner & Henikoff, 2014). We do not see a similar skewed distribution in *B. mori*. Nevertheless, in both organisms, ~50% of the genome is centromere-permissive (Gassmann et al., 2012; Senaratne et al., 2020). However, when we compare the centromere domain sizes, *C. elegans* centromere domains are ~3X smaller than those of *B. mori*. We wanted to assess further whether this difference could be correlated to the gene compaction level in each organism. In other words, are the centromere domains smaller in *C. elegans* simply due to there being less intergenic spaces (i.e. low-nucleosome turnover regions) to which CenH3 could localize to as a result of higher gene density? To test this, we mapped previously published *C. elegans* CenH3-specific genomic sequences from CHIP-seq experiments (Steiner & Henikoff, 2014) to the *C. elegans* genome and after a series of bioinformatic steps to annotate CenH3-enriched “domains”, we calculated the enrichment of centromere domains in *B. mori* over *C. elegans* as a fraction of the domain size (Figure 6). Indeed, this revealed that the *B. mori* genome is overall enriched for larger centromere domains while the *C. elegans* genome is enriched for smaller centromere domains. This comparison gave us a different vantage point to reaffirm that the cues defining holocentromeres are directly linked to chromatin landscape, where larger inter-gene spaces could leave ample space with low chromatin dynamics along which continuous, that is, longer centromere domains can stably form.

Table: Genome and centromere statistics for *C. elegans* and *B. mori*

	<i>C. elegans</i>	<i>B. mori</i>
Genome size	100 mb n=6; 5 autosomes +1X (males) (<i>C. elegans</i> Sequencing Consortium 1998)	460 mb n=28 + 668 unplaced scaffolds (Kawamoto et al., 2019)
# annotated protein coding genes	~20 000 (Wormbase)	~16 000 (Kawamoto et al., 2019)
Gene density per 100 kb	20 (extrapolated from <i>C. elegans</i> Sequencing Consortium 1998.	~3 (Ahola et al., 2014)
% genes: % intergenes	53: 47 (Wikipedia)	36: 64 (Senaratne et al., 2020)
Mean exon length (kb)	147	353

(Suetsugu et al., 2013)		
Mean intron length (kb) (Suetsugu et al., 2013)	410	1904
% TEs in genome	~12 (Suetsugu et al., 2013)	~38 (Kawamoto et al., 2019)
% total repeats in genome	~ 20% (Ahringer & Gasser, 2018)	~46% (Kawamoto et al., 2019)
# coverage by centromere domains	47 % (total) (Gassmann et al., 2012)	54 % (median) (Senaratne et al., 2020)
Median size of centromere domains	10-12 kb (Gassmann et al., 2012)	36 kb (Senaratne et al., 2020)

Figure 6: Enrichment of centromere domains sizes in *B. mori* over *C. elegans*



Finally, it will be useful to combine available centromere profiles for both organisms with additional datasets from techniques like ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) (Buenrostro et al., 2015) and CATCH-IT (Covalent Attachment of Tagged Histones to Capture and Identify Turnover) (Deal et al., 2010), which directly measure chromatin accessibility and nucleosome turnover, respectively and facilitate direct visualization of the link between centromere profiles and chromatin activity.

Evolutionary establishment of holocentric chromosomes in insects

The findings of our study further allow us to propose a molecular mechanism underlying the mono- to holocentric transition in Lepidoptera and possibly other holocentric insects. Ancestral insects were monocentric and CenH3-dependent (Drinnenberg et al., 2014; Melters et al., 2012). CenH3 in ancestral monocentric insects might have self-propagated centromere identity at a restricted locus where it nucleated kinetochore assembly, resembling the monocentromeres in *Drosophila melanogaster* and other monocentric organisms (Barnhart et al., 2011; Black et al., 2004; Carroll et al., 2010; Fachinetti et al., 2013; Guse et al., 2011; Karpen & Allshire, 1997; H. Kato et al., 2013; Logsdon et al., 2015; Mendiburo et al., 2011; Palladino et al., 2020; Roure et al., 2019; Tachiwana et al., 2015). Alternatively, centromeres could have been defined by a genetic mechanism (Kasinathan & Henikoff, 2018). In contrast, in the CenH3-deficient derived state that we characterize in this study, kinetochore assembly and thus, centromere activity occurs chromosome-wide, only antagonized by chromatin disruption processes (Figure 5). Given that CenH3 in insects is essential in monocentric Diptera (M. D. Blower & Karpen, 2001) and its loss is only found in holocentric lineages (Drinnenberg et al., 2014), the establishment of this derived holocentric state is likely to have preceded and subsequently allowed the loss of CenH3. This is also supported by the fact that some holocentric Hemipteran insects have CenH3 homologs (Cortes-Silva et al., 2020). These Hemipterans could therefore represent an intermediate form leading to the establishment of the CenH3-deficient state.

While the molecular function of CenH3 during this possible transition state seen in some Hemipterans is an open question, two events must have occurred to allow the progression to a CenH3-deficient state: (i) centromere identity conferred through an epigenetic loop and/or a genetic component of centromere identity was lost, which led to the establishment of a holocentric architecture; (ii) kinetochore assembly on DNA must have become CenH3-independent, perhaps through the replacement of its ability to attach the kinetochore complex to chromatin by other kinetochore components, such as CENP-T described in this study. Future studies aiming to characterize the centromere architecture of additional holocentric insects will give more resolution to these intermediate events. Furthermore, The fact that meiosis in CenH3-encoding holocentric *C. elegans* proceeds in a CenH3-independent manner (Dumont et al., 2010; Monen et al., 2005), highlights that this organism harnesses the ability to use alternative (CenH3-independent) modes of chromosome segregation. However, CenH3 must be retained in *C. elegans* due to its critical functions at the centromere. The complex CCAN of holocentric insects mediates CenH3- and CENP-C-independent kinetochore assembly by employing alternative means to anchor the

kinetochore (such as via CENP-T) (Cortes-Silva et al., 2020). In contrast, the inner kinetochores of holocentric nematodes lack the CCAN but contain only CenH3 and its direct binding partner CENP-C (Buchwitz et al., 1999; Cheeseman, 2004; Moore & Roth, 2001). Thus, it is possible that CenH3 has a crucial role in *C. elegans* to allow CENP-C to anchor and assemble the outer kinetochore at the centromere by way of the ability of CENP-C to directly interact with CenH3 nucleosomes (Carroll et al., 2010; Cheeseman, 2004; Desai, 2003; Milks et al., 2009; Przewłoka et al., 2011; Screpanti et al., 2011). Thus, retention of CenH3 in holocentric nematodes could be in compensation for the absence of CCAN components other than CENP-C that can anchor and assemble the outer kinetochore- as demonstrated in CCAN-mediated kinetochore assembly in CenH3-deficient holocentric insects (Cortes-Silva et al., 2020)) Future studies to characterize the kinetochores of other CenH3-encoding holocentric organisms will reveal whether other similar cases exist.

A link between holocentromere occurrence and the dosage of CenH3 in independent holocentric organisms has previously been speculated (Cuacos et al., 2015). Such speculation derives from several observations: (I) the presence of two CenH3s in *C. elegans* -HCP-3 and CPAR-1 (CENP-A related 1) (Monen et al., 2005); (ii) CenH3 being essential for *C. elegans* mitosis while is dispensable in male meiosis (Monen et al., 2005); (iii) two CenH3 found in *L. nivea*- Ln CenH3-A and LnCenH3B, where at least LnCenH3B is found at centromeres (Moraes et al., 2011); (iv) entire loss of CenH3 in most holocentric insects (Drinneberg et al., 2014). Holocentric insects provide a good experimental system to test such a correlation given the known existence of Hemipteran species that have lost CenH3 while a few others have retained it (Cortes-Silva et al., 2020). The few CenH3-encoding Hemipteran species can be used to the advantage of understanding the gradation of events leading to holocentromere establishment and CenH3 loss by testing whether their centromeres are already CenH3-independent in spite of having CenH3. Given that these species are known to encode other kinetochore components like CENP-T (Cortes-Silva et al., 2020), in terms of dissecting the transition state, one could profile by ChIP-seq both CenH3 and CENP-T. The two profiles can be expected to overlap only if they are part of the same centromere unit. If they do not overlap, it would indicate a path to loss of CenH3 function at centromeres. On the other hand, dissecting kinetochore assembly in CenH3-encoding Hemiptera will shed light on the critical components needed for kinetochore anchoring and assembly. Collectively, these experiments will give insight into why CenH3 is retained in some holocentric insects.

An intrinsic potential to be holocentric?

Lack of conclusive evidence pointing to reversions to the monocentric form in any eukaryotic lineage is consistent with an advantage over tolerance to DNA damage. This tolerance has been hypothesized within the context of exposure to clastogenic environments (Zedek & Bureš, 2018) and in the specific context of nematode development that is characterized by fixed cell lineages (Pimpinelli & Goday, 1989) (See Introduction). Nevertheless, the large abundance of monocentric organisms over holocentric ones suggests that the challenges associated with acquiring this trait could outweigh its benefits (Mandrioli & Carlo Manicardi, 2012).

Although speculative, theoretically, distinct heterochromatic states found scattered along the chromosome arms of monocentric metazoans (Ho et al., 2014) could provide a stable scaffold for non-specific anchoring of the kinetochore. However, this intrinsic potential to support holocentric morphology might be counteracted by the problems associated with having unrestricted kinetochore activity. One such challenge is the need to overcome merotely. Typically, the kinetochore of each sister chromatid of a chromosome interacts specifically with the microtubules emanating from a single respective spindle pole (and not both). Merotely refers to defects in this bi-polar attachment, where one kinetochore can attach to microtubules emanating from both poles. Studies in several monocentric systems to induce relatively subtle changes in kinetochore size and shape have indeed demonstrated its impacts on maintaining properly attached microtubules. For example, when mammalian cells are treated with anti-mitotic drugs such as Nocadazole, this results in deformed kinetochores that stretch up to around >5 fold and is seen to be accompanied by a drastic increase in lagging chromosomes and aneuploidy (10% as compared to <1-5% in the wild-type condition) that underlie merotelically attached sister kinetochores (Cimini et al., 2001). Thus, intuitively, in holocentric organisms it would seem that the proneness to merotely would be drastically increased in the presence of naturally stretched kinetochores. Yet, being holocentric is not described to impart such an elevated risk. Although it is still unclear what the underlying mechanisms might be, efforts to understand kinetochore orientation in the holocentric condition have led to bits of evidence from *C. elegans* that give some insight to the factors associated with maintaining stable bi-polar attachment. For example, screening for chromosome segregation mutants in *C. elegans* led to the identification of a conserved subunit of the condensin-2 complex subunit, HCP-6 (holocentric protein 6), that was found to have critical roles for ensuring the correct alignment, condensation and maintenance of rigidity of mitotic chromosomes to prevent chromosome twisting and to ensure bi-polar attachment of sister kinetochores (Stear & Roth, 2002), thus making a clear link

in how defects in chromosome condensation translates to chromosome mis-segregation by way of merotely.

Previous studies also in *C. elegans* revealed that the pattern of condensin 1 & 2 along holocentric mitotic chromosomes is drastically different from the monocentric form (Csankovszki et al., 2009). While in the monocentric form condensin 1 & 2 show an alternating distribution within the core of each sister chromatid arm with an enrichment of condensin 2 at the centromere, in the holocentric *C. elegans*, condensin 2 is restricted to the polar surface of sister chromatids where the extended centromere lies, while condensin 1 is broadly distributed internal to condensin 2 (reviewed by (Csankovszki et al., 2009). This lends to the idea that a differential distribution of condensin 1 & 2 between the mono- and holocentric forms may indeed play a role in shaping the architecture and placement of each chromosome morphology (Csankovszki et al., 2009). While it is likely that kinetochore orientation is regulated by similar mechanisms in both mono- and holocentric organisms (Dernburg, 2001), it would indeed be worthwhile to investigate whether a particular organization of condensin regulates the propensity to merotely in holocentric organisms using additional genetically tractable holocentric systems like *B. mori*, and further, to compare that to any organism having a large centromere such as the regional monocentromere of human. One additionally interesting hypothesis regarding the adaptation of holocentric chromosomes to prevent merotely is that the size of holocentric chromosomes could be optimal for allowing bi-polar orientation (Dernburg, 2001). Here, Dernburg pointed out that the size of *C. elegans* chromosomes (~14 to 21 mb) lies within the separating distance (~20 mb) at which dicentromeres of mammalian chromosomes are implied to act in concert such that dicentric chromosomes could remain mitotically stable (Dernburg, 2001). Thus, this hypothesis essentially implies that all kinetochores along a holocentric chromatid are within a distance that allows them to act in concert as a single unit in order to specifically bind microtubules radiating from a single pole.

Finally, in contrast to the *in vivo* scenario where the presence of centromere-specific components like CenH3 would mediate the localization of the kinetochore to a specified centromere, testing the mitotic potency of extrachromosomal DNA in a CenH3-independent environment such as in a lepidopteran system, would alleviate a prior notion of a restriction to form a centromere at a specific site. This can be used to test the flexibility of a piece of DNA to build an alternative chromatin template that can be used to then build a centromere. While these studies have been carried out in holocentric systems in the past (Stinchcomb et al., 1985), where extrachromosomal DNA arrays in *C. elegans* are shown to recruit centromeric chromatin and kinetochore proteins to remain mitotically potent (Wong et al., 2020; Yuen et al., 2011), similar experiments in a naturally CenH3-deficient holocentric system would have overarching implications on how centromere regulation is perceived at the very basic

level across eukaryotes. Ongoing experiments in our lab should shed more light on this aspect in the future.

Methods

Lepidopteran cell lines and culture conditions

Cultured silkworm ovary-derived BmN4 (ATCC catalog # CRL-8910; RRID: CVCL_Z633), BmN4-SID-1 (systemic RNA interference-deficient 1) (RRID: CVCL_Z091) (I. Kobayashi et al., 2012) and *T. ni* Hi5 cell lines (Granados et al., 1986) were maintained in Sf-900 II SFM medium (GIBCO catalog # 10902-088) supplemented with (BmN4, BmN4-SID-1) or without (Hi5) 10% FBS (fetal bovine serum) (Eurobio catalog # CVFSVF0001), antibiotic-antimycotic (GIBCO catalog # 15240-062) and 2 mM (millimolar) L-glutamine (GIBCO catalog # 25030-024) at 27 °C (degrees Celsius).

Construction of a stable cell line expressing 3xFLAG-tagged *B. mori* H3.3

3xFLAG N-terminally tagged H3.3 c (complementary) DNA derived from *B. mori* was cloned into pIZV5 plasmid vector (Invitrogen, catalog # V800001) using Kpn1 and Xho1 restriction enzymes. To construct the stable BmN4 cell line around 1-5 µg (micrograms) of plasmid DNA was transfected into 10⁶ BmN4 cells using Cellfectin II (GIBCO, catalog # 10362100) according to the manufacturer's instructions and was selected using 300 µg/ml (micrograms/milliliter) Zeocin (GIBCO, catalog # R25001) until no viable untransfected cells were observed. Expression of the H3.3 transgene was confirmed by IF.

Validation of *B. mori* CENP-T mouse antibody specificity by RNAi

BmN4-SID-1 cells were grown in 24-well plates for 3 days with or without 400 pg/µl (picograms/microliter) double-stranded RNA targeting CENP-T. At the end of 3 days, RNAi-treated or WT cells were fixed with 100% ice-cold MeOH (methanol) for 10 min at -20 °C and then processed for IF microscopy.

Preparation of mitotic chromosome spreads

BmN4 cells were grown in 6-well plates and collected by mild centrifugation (300 g (units of gravity)) at room temperature. Supernatant containing growth medium was decanted and cell pellet was gently re-suspended in hypotonic buffer (80% water, 20% PBS (phosphate buffered saline)) that was added drop-wise under mild vortex to the cell pellet. Cells were incubated in hypotonic buffer for 30 min at room temperature. Swollen cells in hypotonic buffer were then aliquoted into disposable funnels and cytopun onto coverslips using a Shandon cytopspin 3 centrifuge. Chromosome spreads were un-mounted from cytopspin and were fixed immediately with 4% PFA (paraformaldehyde) for 10 min at room temperature and processed for IF microscopy.

IF

Cells fixed in either 100% ice-cold MeOH (stainings in BmN4-SID-1 cells) or 4% PFA (mitotic chromosome spreads and H3K9me3 stainings in BmN4 cells and mouse ES (embryonic stem) cells) were permeabilized using 0.3% Triton x-100 in PBS. Cells were then blocked in 3% BSA (Bovine Serum Albumin)-PBS. Primary antibody incubations were done in blocking buffer overnight at 4 °C. The following primary antibodies were used at 1:1000 dilution: anti-CENP-T serum (rabbit or mouse polyclonal), anti-Dsn1 serum (rabbit polyclonal) generated by Covalab (Villeurbanne, FR) (Cortes-Silva et al., 2020), and H3K9me3 antibodies: rabbit polyclonal, Abcam, ab8898 and mouse monoclonal, MBL MABI0318. The next day, cells were washed three times with 0.3% Triton x- 100 in PBS and were incubated for 1 hour at 4 °C with secondary antibodies diluted to 1:1000 in blocking buffer. The following fluorescent-conjugated secondary antibodies were used: goat anti-rabbit IgG (Immunoglobulin G) Alexa Fluor 488 (Thermo Fisher Scientific, catalog # A-11034, RRID AB_2576217), goat anti-rabbit IgG Alexa Fluor 568 (Thermo Fisher Scientific, A-11011, RRID AB_143157), and goat anti-mouse IgG Alexa Fluor 568 (Thermo Fisher Scientific, catalog # A-11004, RRID AB_2534072). Cells were washed three times with 0.3% Triton x-100 in PBS and counterstained with DAPI for 3 min at room temperature (Sigma catalog # D9542) before washing again in 0.3% Triton x-100 in PBS and mounting samples in Vectashield Antifade Mounting Medium (Vector Laboratories catalog # H-1000; RRID:AB_2336789).

Microscopy

Images of chromosome spreads and H3K9me3 stainings in mitotic or interphase BmN4 or mouse ES cells were acquired on a LSM780 confocal microscope. Z stacks were acquired at 0.1-0.2 μm intervals using the 100X oil objective. Images of BmN4-SID-1 interphase cells stained for CENP-T in WT or RNAi conditions were acquired on a Zeiss Axiovert Z1 light microscope. Z stacks were acquired at 0.2 μm intervals using the 100X oil objective.

Quantification of fluorescence intensity was performed using the Fiji software (Schindelin et al., 2012) on unprocessed TIFF (Tagged Image File Format) images. IF signal in interphase BmN4-SID-1 cells stained for either CENP-T or Alexa Fluor 568 were quantified. Nuclei of 10 cells were manually segmented using DAPI signal. The mean fluorescence intensity of each nucleus was then measured and corrected for background. For background correction, the average of mean intensities of three random circular regions of fixed size (10x10 pixels) placed outside the nuclear areas was determined and subtracted from the CENP-T or Alexa-Fluor-specific IF signal of each nucleus.

X-ChIP using in-house protocol

X-ChIP was performed as previously described (Skene & Henikoff, 2015) with the following modifications. Two confluent T75 flasks (Thermofisher, catalog # 156499) of BmN4 cells (for CENP-T, Dsn1, RNA Pol II and FLAG ChIPs) or Hi5 cells (for CENP-T ChIP) were used. Cells were cross-linked in freshly prepared 1% MeOH-free formaldehyde (Thermofisher catalog # 28906) for 10 min at room temperature. Cross-linking was quenched by adding glycine to 125 mM for 2 min at room temperature. Cells were then washed in ice-cold PBS and incubated for 10 min with 150 μ l ice-cold lysis buffer (1% SDS (sodium dodecyl sulfate), 10 mM EDTA (Ethylenediaminetetraacetic acid), 50 mM Tris-HCl (Tris hydrochloride) pH 8.1) with cComplete Protease Inhibitor Cocktail (Roche catalog # 11697498001). To the cell lysates, 1350 μ l ChIP buffer (1% Triton X-100, 150 mM NaCl (Sodium Chloride), 2 mM EDTA, 20 mM Tris-HCl pH 8.1,) with Protease inhibitor cocktail was added along with 4.5 μ l CaCl₂ (Calcium Chloride) 1 M (molar) (3 mM final) and then pre-warmed for 2 min at 37 °C. Nuclei were then treated with 1 or 2 units of MNase (Sigma catalog # N3755-500UN) for 15 min (CENP-T and FLAG ChIPs), 30 min (CENP-T ChIP), 45 min (CENP-T, Dsn1, RNA Pol II ChIPs) or 60 min (CENP-T ChIP), respectively at 37 °C. MNase reaction was stopped by adding a mix of 30 μ l EDTA (0.5 M stock) and 60 μ l of EGTA (ethylene glycol-bis(β -aminoethyl ether)-N,N,N',N'-tetraacetic acid (0.5 M stock). Each MNase-treated nuclei sample was then sonicated using a Covaris E220 sonicator under the following parameters: 150 sec (WT CENP-T ChIPs in BmN4 cells) or 250 sec (Dsn1, RNA Pol II and 20E CENP-T ChIPs in BmN4 cells; CENP-T ChIP in Hi5 cells), Duty 10%, Power 75 W (watts), cycles/burst 200, 7 °C. Sonicated chromatin was centrifuged 3 min at 16000 g and clear supernatant containing the solubilized chromatin was saved either as input or for ChIP. Anti-CENP-T serum (rabbit polyclonal), anti-Dsn1 serum (rabbit polyclonal) or anti RNA Pol II antibody (rabbit polyclonal, Abcam, ab817) diluted in 0.3% Triton X-100 or 0.02% tween-20 was incubated with Protein A dynabeads (Thermofisher, catalog # 10001D) for 10 min at room temperature to allow for Protein A-beads-antibody binding. Antibody-bound beads were washed with ChIP buffer and mixed with input chromatin. Alternatively, commercial beads-anti-FLAG M2 antibody (Sigma catalog # M8823; RRID:AB_2637089) was directly added to input chromatin for control FLAG X-ChIP. All samples were incubated overnight at 4 °C. Chromatin-bound beads were collected the next day on a magnetic rack and washed with the following ice-cold buffers: once with low-salt TSE 1 (0.1% SDS, 1% Triton X-100, 150 mM NaCl, 2 mM EDTA, 20 mM Tris-HCl pH 8.1.); four times with high-salt TSE II (0.1% SDS, 1% Triton X-100, 500 mM NaCl, 2 mM EDTA, 20 mM Tris-HCl pH 8.1); and three times with 1x TE (Tris-EDTA) buffer. DNA was directly extracted from chromatin-beads or input by adding DNA extraction buffer (20 mM Tris-HCl pH 8.1, 10 mM EDTA, 5 mM EGTA, 300 mM NaCl, 1% SDS) and incubating at 37 °C followed by reversing cross-links by addition of proteinase K (Qiagen, catalog #

19131) and incubation overnight at 65 °C. DNA was isolated with Phenol: Chloroform extraction and precipitated with NaOAc (Sodium Acetate) and 100% EtOH (ethanol) in the presence of glycogen. DNA was finally re-suspended in 1x TE containing RNase (1 µg/µl) and incubated for 15 min at 37 °C. Nucleosome profiles for Input and ChIP DNA were analyzed using a Agilent bioanalyzer or Agilent 4200 TapeStation with a DNA high sensitivity kit.

X-ChIP using commercial protocol

ChIP and DNA extraction was performed as described in the Diagenode iDeal ChIP-seq kit for histones (Diagenode, catalog # C01010051/ C01010057) using two confluent T75 flasks of BmN4 cells. Chromatin was sheared using a Covaris E220 sonicator. Solubilized chromatin was incubated overnight at 4 °C with the following antibodies: H3K4me3 (positive control ChIP-seq grade antibody provided in Diagenode iDeal ChIP-seq kit); H3K9me2 (mouse monoclonal, MBL, MABI0317; and rabbit polyclonal, Diagenode, C15410060); H3K9me3 (mouse monoclonal, MBL, MABI0318; and rabbit polyclonal, Abcam, ab8898); H3K27me3 (Rabbit polyclonal, Cell Signaling Technology, C36B11); and H3K36me3 (rabbit polyclonal, Abcam, ab9050). Nucleosome profiles for Input and ChIP were analyzed using a Agilent bioanalyzer or Agilent 4200 TapeStation with a DNA high sensitivity kit.

Native ChIPs

ChIP and DNA extraction were performed as described (Orsi et al., 2015) using two confluent T75 flasks of BmN4 or Hi5 cells. Solubilized chromatin was incubated overnight at 4 °C with the following antibodies: H3K27me3 (Rabbit polyclonal, Cell Signaling Technology, C36B11); H3K36me3 (rabbit polyclonal, Abcam, ab9050); H3K9me2 (rabbit polyclonal, Activ Motif, AM39753); H3K9me3 (rabbit polyclonal, Abcam, ab8898); H3 (rabbit polyclonal, Abcam, ab1791); for 3X-FLAG H3.3 ChIPs: anti-FLAG M2 antibody (mouse monoclonal, Sigma, catalog # F1804; RRID: AB_262044). Nucleosome profiles for Input and ChIP were analyzed using a Agilent bioanalyzer or Agilent 4200 TapeStation with a DNA high sensitivity kit.

Next-generation sequencing and ChIP-seq data analysis

All steps of Illumina library preparation and sequencing were carried out at the Curie Institute's sequencing platform. Adapter trimmed, single-end Illumina reads of 100 bp length were mapped using Bowtie2 (Langmead & Salzberg, 2012) with default parameters to the *B. mori* genome assembly (Kawamoto et al., 2019) downloaded from Silkbase: <http://silkbases.ab.a.u-tokyo.ac.jp>, which was modified to extract only assembled chromosomes 1 to 28 or the *T. ni* genome (Fu et al., 2018) downloaded

from the Cabbage Looper Database: <https://cabbagelooper.org>. After removal of duplicates using Picard tools (<http://broadinstitute.github.io/picard/>), Deeptools bamCompare function (Ramírez et al., 2016) was used to generate ChIP-seq signal tracks represented as histograms of the average log₂-ratio of RPKM (reads per kilobase per million mapped reads) -normalized read counts in IP over Input in genome-wide 1 kb windows that were visualized in IGV (Integrative Genomics Viewer) (Robinson, 2011). Deeptools multiBigWigSummary function (Ramírez et al., 2016) was used to compute average log₂-ratio of RPKM-normalized read count in IP over Input in genome-wide 10 kb windows for making scatterplots of the correlation between different ChIP-seq targets. Scatterplots and Pearson correlation (*r*) calculations were done in RStudio (RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>) after filtering out those 10 kb windows with zero mapped reads in both IP and Input.

Annotation of CENP-T domains

CENP-T ChIP-seq signal originally in 1 kb windows was averaged over 2 kb. The averaged windows with positive scores (log₂-ratio>0) within a genomic distance of 5 kb were merged using BEDTools (Quinlan & Hall, 2010). This cut-off for merging distance was determined so that the likelihood of finding consecutive 1 kb windows with positive CENP-T signal was significant at $p \leq 0.05$. Average CENP-T ChIP-seq signal was re-computed over the merged coordinates after removing any merged intervals of size <5 kb. Any merged intervals with overall negative ChIP-seq scores were further removed. These were defined as positive domains. Negative domains were similarly annotated using a reciprocal approach where log₂ scores ≤ 0 were considered as negative scores. BEDTools (Quinlan & Hall, 2010) was used to subtract negative domains from positive domains to extract final CENP-T domains.

Repeat analyses using RepeatMasker

The *B. mori* genome assembly (chromosomes 1 to 28) and newly-annotated CENP-T domains were searched using RepeatMasker software version 4.08 and RMBlast version 2.10.0+ (<http://www.repeatmasker.org>) with a custom library of consensus transposon sequences for *B. mori* (kind gift from the Pillai Lab, University of Geneva). Simple repeats and low-complexity repeats accounting for <1% of the genome (Kawamoto et al., 2019) were omitted from the analyses. The percentage of interspersed repeats in CENP-T domains and genome-wide was calculated as the total number of base pairs covered by interspersed repeats as a fraction of the total genome size.

Repeats analyses using k-mer clustering

k-mer based repeat analysis pipeline from the Straight Lab (Smith et al., 2020): Single-end Illumina reads (adapter- trimmed, PCR (polymerase chain reaction) duplicates removed) for CENP-T ChIP-seq and Input were used to generate k-mer databases for each dataset at k-mer lengths of 10 bp and 25 bp. The abundance of each k-mer in both datasets was counted and normalized to the total number of bp in that dataset. K-mers found fewer than 10 times in either dataset were excluded from the analysis. Normalized k-mer counts in the CENP-T dataset (y-axis) were plotted as a function of normalized k-mer counts in input (x-axis). Enrichment values for each k-mer were calculated as the ratio of normalized count in CENP-T dataset over the normalized count in input dataset. Different enrichment cut-offs were defined according to the number of median absolute deviations away from the median enrichment ratio.

Transcriptional profiling

Total RNA was isolated from one confluent T75 flask of BmN4 or Hi5 cells using Trizol reagent (Invitrogen, catalog # 15596018) following the manufacturer's instructions. PolyA-selected RNA-seq libraries were prepared using the Illumina TruSeq stranded mRNA protocol and sequenced at the Curie Institute's sequencing platform. Adapter-trimmed, paired-end reads of 100 bp length were mapped to the *B. mori* genome assembly (chromosomes 1 to 28) using STAR (Spliced Transcripts Alignment to a Reference) version 2.7 (Dobin et al., 2013). The number of mapped RNA-seq reads in annotated genes (downloaded from SilkBase: <http://silkbase.ab.a.u-tokyo.ac.jp>) or in genome-wide 10 kb windows were counted using HTSeq software (Anders et al., 2015). Read counts were log₂ transformed into TPM scores to evaluate normalized expression levels. 1 kb resolution TPM-normalized RNA-seq coverage tracks were further generated using Deeptools (Ramírez et al., 2016) BamCoverage function for visualization in IGV (Robinson, 2011) as histograms.

20E treatment and washout

20E-treated or DMSO-control samples were prepared by growing BmN4 cells in 5 µg/ml of 20E-hydroxyecdysone (Sigma-Aldrich, H5142) or <1% DMSO for 48 hours, following which RNA-seq (20E- treated and DMSO-treated) or CENP-T X-ChIP-seq (20E-treated) was performed. For 20E-washout RNA- seq and ChIP-seq experiments, cell layers growing for two days in the presence of 20E were softly rinsed with fresh growth medium thrice to dilute out the hormone without disrupting the monolayer. Cell layers were then allowed to grow in conditioned growth medium* for a further 5

days (RNA-seq) or 10 days (ChIP-seq) before harvesting. Cells growing in conditioned medium were expanded as appropriate by splitting with 50% fresh growth medium and 50% conditioned medium. * Conditioned medium was collected from WT BmN4 cells growing for the same time period as the 20E treatment (2 days). Medium was decanted into a falcon tube and centrifuged at 1000 g for 10 min to remove any cells in the medium. Supernatant was filtered twice through a 0.22 μ m filter and stored at 4 °C for a maximum of 3 hours until use.

Criteria for identifying differentially-expressed genes from 20E- or DMSO-treated RNA-seq datasets

Up-regulated genes were defined as those whose expression increased by at least 5-fold and had minimum expression levels of at least 5 log₂ TPM units after 20E treatment. Down-regulated genes were defined as those whose expression reduced by at least 5-fold in the 20E-treated condition and had minimum starting expression levels of at least 5 log₂ TPM units in the DMSO control. Functional annotations of *B. mori* gene models were a kind gift from M. Kawamoto, University of Tokyo.

Criteria for identifying CENP-T-enriched or-depleted genomic windows

Genomic windows from 20E-treated or -washout datasets with CENP-T level differences that are equal to or exceed 3 standard deviations subtracted or added to the mean log₂ enrichment score of all bins in the WT were selected. In addition, only those 10 kb windows with total loss or gain of CENP-T from a previously enriched or depleted state, respectively, were considered for our analyses.

Identification of orthologous genes

B. mori and *T. ni* proteome datasets from Silkbase (<http://silkbases.ab.a.u-tokyo.ac.jp>, published in (Kawamoto et al., 2019)) and Cabbage Looper Database (<https://cabbagelooper.org>, published in (Fu et al., 2018)) were used. Orthologous genes were selected as those identified as reciprocal best hits in blastp (protein-protein basic local alignment search tool) searches (Altschul et al., 1990) of *B. mori* against *T. ni* proteome and *vice versa*. These analyses revealed 10212 orthologs between the two organisms. We filtered out any orthologs that had multiple hits in the reciprocal organism. This left us with a total of 9533 orthologs for differential expression analyses.

Criteria for identifying differentially-expressed orthologous genes

Highly expressed genes for either *B. mori* or *T. ni* were defined as those that had a difference in expression level of at least 5 units to which was added (for *B. mori*) or subtracted (for *T. ni*) the median difference in expression of all 9533 orthologs. Difference in expression was calculated as $\log_2 \text{TPM}+1(B. mori - T. ni)$ for each gene.

Quantification and statistical analysis

Statistical details of experiments are detailed in the figure legends. Statistical analyses were performed in RStudio (RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>).

Supplementary Figures

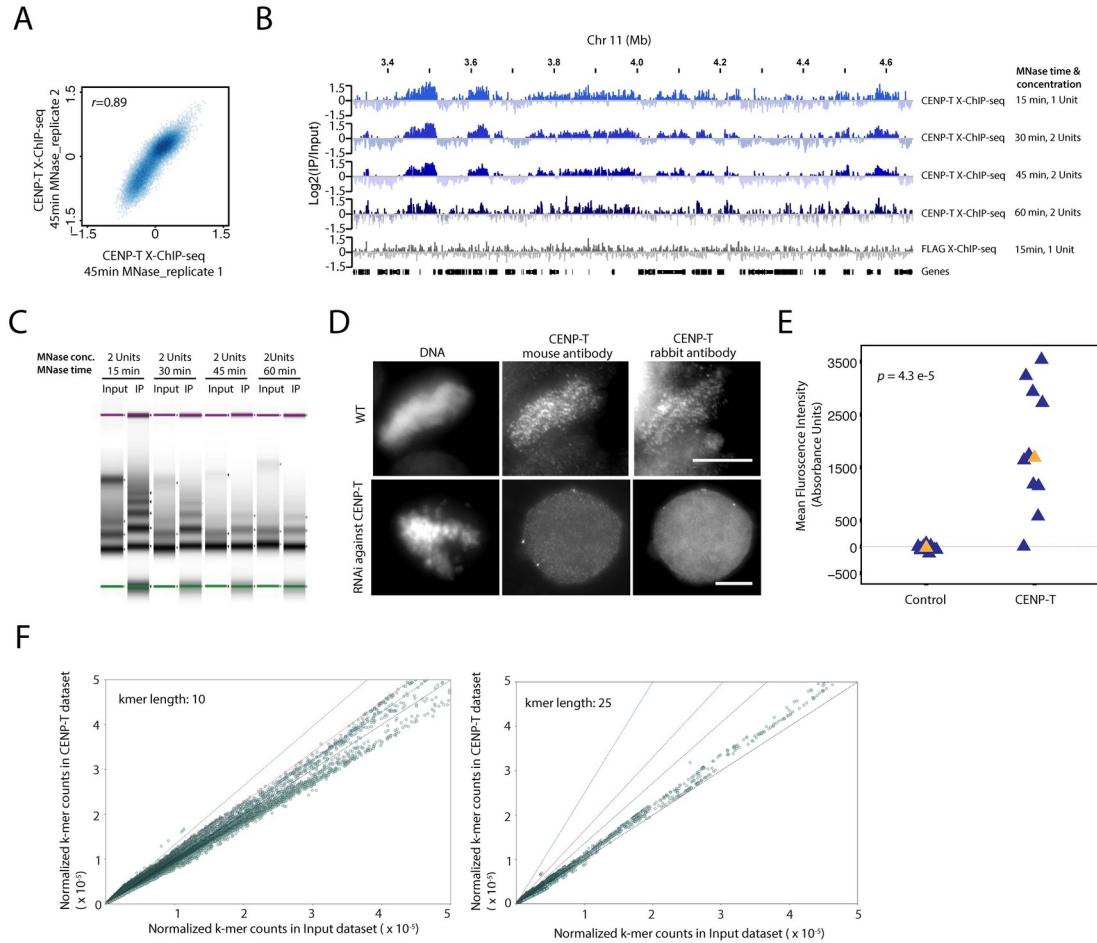


Figure S1: CENP-T localizes to non-repetitive domains during interphase and mitosis

A) X-ChIP-seq profiles are well-correlated across varied MNase conditions. Genome-wide correlation plot of CENP-T occupancy from replicate X-ChIP-seq profiles generated with 2 units of MNase and 45 min of digestion. Average log₂ ratios of IP/Input in genome-wide 10 kb windows were used for plotting and calculating the Pearson correlation coefficient (r). **B)** Genome-browser snapshot of a representative portion of *B. mori* chromosome 11 for CENP-T X-ChIP-seq at varied MNase conditions, FLAG X-ChIP-seq (negative control), and annotated genes. Time and concentration of MNase are annotated adjacent to each track. ChIP-seq signal is represented as the average log₂ ratio of IP/Input in genome-wide 1 kb windows. **C)** Agarose gel image of nucleosome enrichment profiles for input and CENP-T pull-down at 15 min, 30 min, 45 min, and 60 min of MNase digestion (2 units), respectively. IPs from chromatin subject to 30 min and 45 min of MNase digestion from the same experiment were sequenced and resulting X-ChIP-seq profiles are shown above in Figure S1B tracks 2 and 3, respectively. **D)** Validation of CENP-T antibody specificity. Representative images of BmN4-SID-1 mitotic cells in either WT condition (top panel) or after three days of RNAi treatment targeting CENP-T (bottom panel). Cells were stained for DNA (left), pre-validated CENP-T antibody generated in mouse host used in this study for co-stainings with Dsn1 (middle) and CENP-T antibody generated in rabbit host (right) (Cortes-Silva et al., 2020). Scale bar: 5 μ m. **E)** CENP-T is present in interphase. Quantifications of mean fluorescence intensity of CENP-

T as compared to control cells stained for secondary Alexa Fluor 568 antibody in BmN4-SID-1 interphase cells (n=10 cells). Statistical significance was tested using the Wilcoxon-Mann-Whitney test. Median is indicated with an orange triangle. **F)** CENP-T domains are non-repetitive. k-mer plots showing the normalized k-mer counts (green circles) in CENP-T X-ChIP-seq vs Input datasets for k-mer lengths of 10 bp (left) and 25 bp (right), respectively. Black, blue, green and red diagonal lines depict increasingly stringent k-mer enrichment ratio cut-offs.

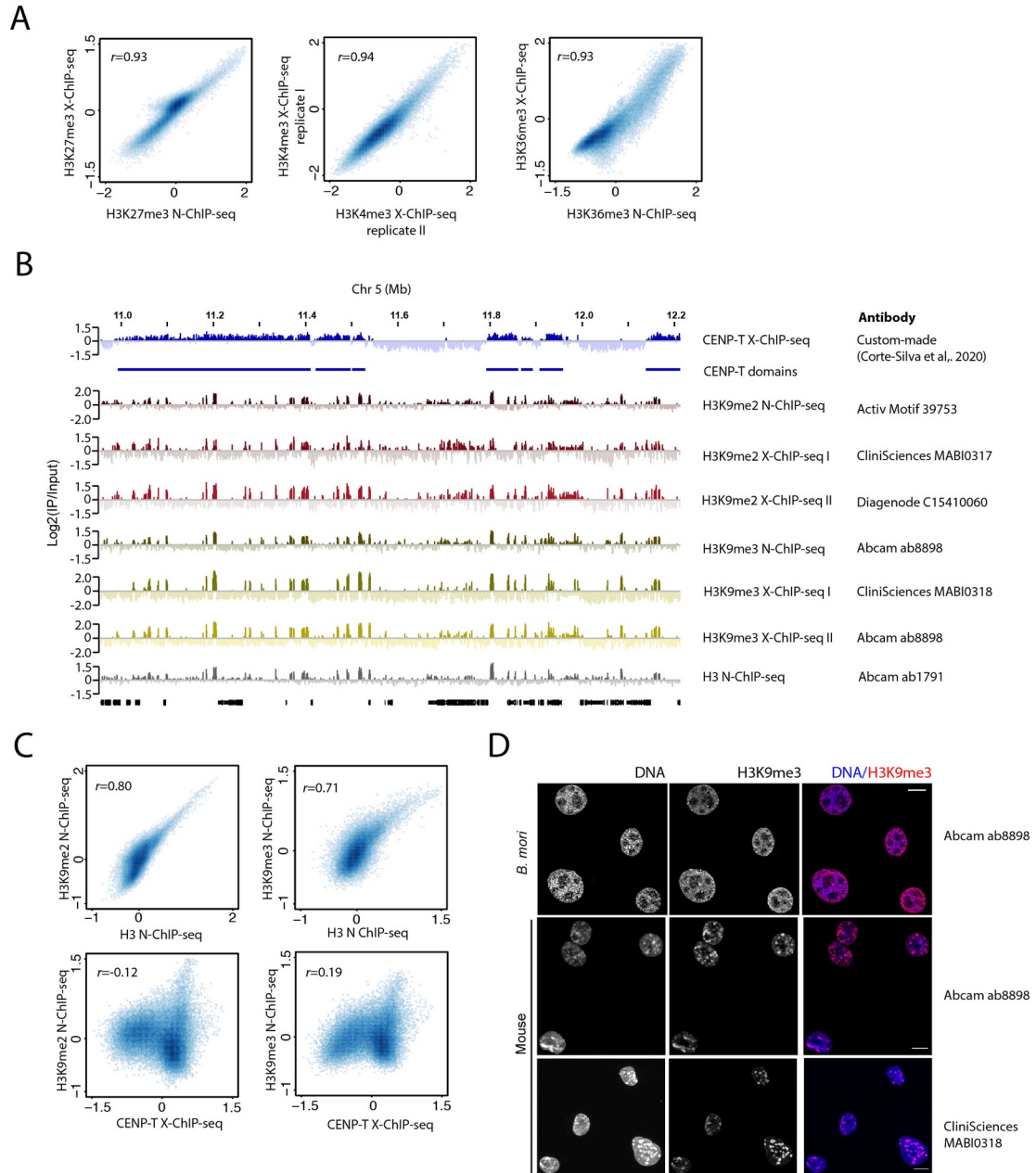


Figure S2: Histone mark patterns in *B. mori*

A) Genome-wide correlation plots of H3K27me3 N-ChIP-seq vs H3K27me3 X-ChIP-seq (Left); replicates of H3K4me3 X-ChIP-seq (middle); and H3K36me3 N-ChIP-seq vs H3K36me3 X-ChIP-seq (right). Average log₂ ratios of IP/Input in 10 kb windows were used for plotting and calculating the pearson correlation coefficient (*r*). **B)** Genome browser snapshot of a representative portion of *B. mori* chromosome 5 for CENP-T X-ChIP-seq, H3K9me2 N-ChIP-seq, H3K9me2 X-ChIP-seq I and II, H3K9me3 N-ChIP-seq, H3K9me3 X-ChIP-seq I and II, H3 N-ChIP-seq and annotated genes. The antibody used for each H3K9 IP is annotated adjacent to the respective track. ChIP-seq signal is represented as the average log₂ ratio of IP/Input in genome-wide 1 kb windows. **C)** Genome-wide correlation plots of

H3K9me2 N-ChIP-seq or H3K9me3 N-ChIP-seq vs H3 N-ChIP-seq (top) and vs CENP-T X-ChIP-seq (bottom). Average log₂ ratios of IP/Input in 10 kb windows were used for plotting and calculating the pearson correlation coefficient (*r*). **D)** Representative IF microscopy images of *B. mori* and mouse interphase cells stained for H3K9me3 (red) and DNA (blue). The anti-H3K9me3 antibodies that were used are annotated to the right. Chromocenters constituting H3K9me3-enriched heterochromatin can be seen as intensely DAPI- and H3K9me3-stained nuclear foci in the mouse cells for both antibodies, whereas homogenous DAPI staining can be seen for *B. mori*. Scale bar: 10µm.

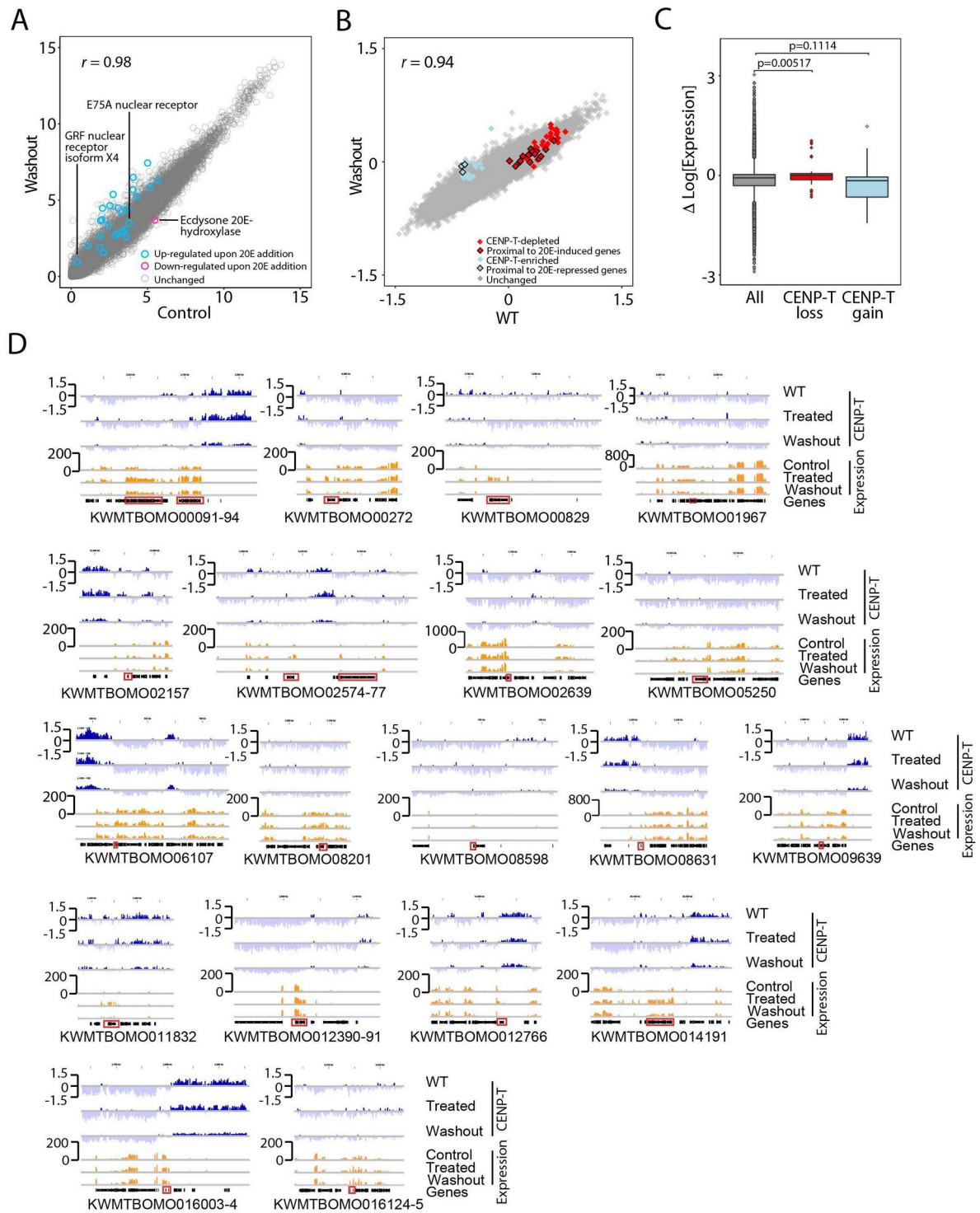


Figure S3: 20E-washout restores transcription and CENP-T occupancy levels.

A) Genome-wide correlation plot comparing the expression levels of annotated genes in 20E-washout vs DMSO-control. Pre-identified subset of twenty-six differentially-expressed genes upon 20E treatment are marked as pink and blue circles, respectively. Functions of two up-regulated genes and one down-regulated that were linked to three cases of differential CENP-

T occupancy in the 20E-treated condition are annotated in the plot. Log₂-transformed TPM is used to represent expression level per gene and to calculate the pearson correlation coefficient (r). **B)** Genome-wide correlation plot comparing CENP-T occupancy in 20E-washout vs WT conditions in 10 kb windows (grey boxes). Pre-identified 10 kb windows with differential CENP-T occupancy upon 20E-treatment are marked as red or blue filled boxes, respectively. Average log₂ ratios of IP/Input in 10 kb windows were used for plotting and calculating the pearson correlation coefficient (r). **C)** Boxplot: difference in expression levels between 20E-washout and DMSO-control in genome-wide 10 kb windows (grey) and in the subset of pre-identified 10 kb windows with depleted or enriched CENP-T occupancy after 20E treatment (red and blue, respectively). Difference in expression was calculated for each 10 kb window by subtracting the log₂ TPM score of control from 20E-washout. Statistical significance was tested using the Kolmogorov-Smirnov test. **D)** Genome browser snapshots of *B. mori* chromosomes for CENP-T X-ChIP-seq (blue) and RNA-seq (orange) surrounding twenty-three remaining up-regulated genes (annotated) that were identified with user-defined expression cut-offs. Tracks: CENP-T X-ChIP-seq profiles (blue) and RNA-seq profiles (orange) for WT (DMSO control for RNA-seq), 20E-treated and 20E-washout. ChIP-seq signal is represented as the average log₂ ratio of IP/Input in genome-wide 1 kb windows. RNA-seq signal is represented as log₂ transformed BPM.

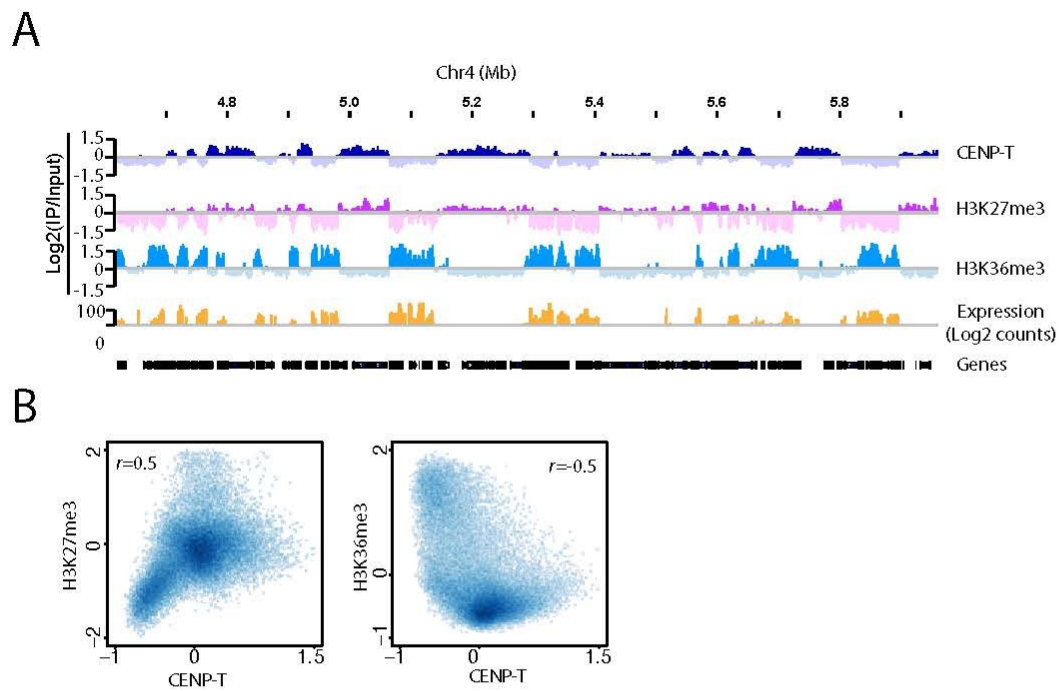


Figure S4: Centromere specification is conserved among CenH3-lacking Lepidoptera.

A) Genome-browser snapshot of a representative portion of *T. ni* chromosome 4 for CENP-T X-ChIP-seq, H3K27me3 N-ChIP-seq, H3K36me3 N-ChIP-seq and annotated genes. ChIP-seq signal is represented as the average log₂ ratio of IP/Input in genome-wide 1 kb windows. CENP-T ChIP in *T. ni* was carried out using the same validated CENP-T antibody used for *B. mori* ChIPs. **B)** Genome-wide correlation plots of *T. ni* CENP-T occupancy and H3K27me3 (left); and H3K36me3 (right). Average log₂ ratios of IP/Input in 10 kb windows were used for plotting and calculating the Pearson correlation coefficient (*r*).

Bibliography

- Ahmad, K., & Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Molecular Cell*, 9(6), 1191–1200.
- Ahola, V., Lehtonen, R., Somervuo, P., Salmela, L., Koskinen, P., Rastas, P., Välimäki, N., Paulin, L., Kvist, J., Wahlberg, N., Tanskanen, J., Hornett, E. A., Ferguson, L. C., Luo, S., Cao, Z., de Jong, M. A., Duploux, A., Smolander, O.-P., Vogel, H., ... Hanski, I. (2014). The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nature Communications*, 5(1), 4737. <https://doi.org/10.1038/ncomms5737>
- Ahringer, J., & Gasser, S. M. (2018). Repressive Chromatin in *Caenorhabditis elegans*: Establishment, Composition, and Function. *Genetics*, 208(2), 491–511. <https://doi.org/10.1534/genetics.117.300386>
- Akera, T., Chenoweth, D. M., Janke, C., Schultz, R. M., & Lampson, M. A. (2017). *Spindle asymmetry drives non-Mendelian chromosome segregation*. 6.
- Akiyoshi, B., & Gull, K. (2014). Discovery of Unconventional Kinetochores in Kinetoplastids. *Cell*, 156(6), 1247–1258. <https://doi.org/10.1016/j.cell.2014.01.049>
- Albertson, D. G., & Thomson, J. N. (1982). The kinetochores of *Caenorhabditis elegans*. *Chromosoma*, 86(3), 409–428. <https://doi.org/10.1007/BF00292267>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 8.
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Barnhart, M. C., Kuich, P. H. J. L., Stellfox, M. E., Ward, J. A., Bassett, E. A., Black, B. E., & Foltz, D. R. (2011). HJURP is a CENP-A chromatin assembly factor

- sufficient to form a functional de novo kinetochore. *The Journal of Cell Biology*, 194(2), 229–243. <https://doi.org/10.1083/jcb.201012017>
- Black, B. E., Foltz, D. R., Chakravarthy, S., Luger, K., Woods, V. L., & Cleveland, D. W. (2004). Structural determinants for generating centromeric chromatin. *Nature*, 430(6999), 578–582. <https://doi.org/10.1038/nature02766>
- Blower, M. D., & Karpen, G. H. (2001). The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nature Cell Biology*, 3(8), 730–739. <https://doi.org/10.1038/35087045>
- Blower, Michael D., Sullivan, B. A., & Karpen, G. H. (2002). Conserved Organization of Centromeric Chromatin in Flies and Humans. *Developmental Cell*, 2(3), 319–330. [https://doi.org/10.1016/S1534-5807\(02\)00135-1](https://doi.org/10.1016/S1534-5807(02)00135-1)
- Bodor, D. L., Valente, L. P., Mata, J. F., Black, B. E., & Jansen, L. E. T. (2013). Assembly in G1 phase and long-term stability are unique intrinsic features of CENP-A nucleosomes. *Molecular Biology of the Cell*, 24(7), 923–932. <https://doi.org/10.1091/mbc.e13-01-0034>
- Boveri, Theodor. (1888). *Die Befruchtung und Teilung des Eies von Ascaris megalcephala* (Vol. 2). Fischer.
- Brown, S. W. (1966). Heterochromatin. *Science (New York, N.Y.)*, 151(3709), 417–425.
- Brown, W. R. A., Thomas, G., Lee, N. C. O., Blythe, M., Liti, G., Warringer, J., & Loose, M. W. (2014). Kinetochore assembly and heterochromatin formation occur autonomously in *Schizosaccharomyces pombe*. *Proceedings of the National Academy of Sciences*, 111(5), 1903–1908. <https://doi.org/10.1073/pnas.1216934111>
- Buchwitz, B. J., Ahmad, K., Moore, L. L., Roth, M. B., & Henikoff, S. (1999). A histone-H3-like protein in *C. elegans*. *Nature*, 401(6753), 547–548. <https://doi.org/10.1038/44062>

- Buck, R. C. (1967). Mitosis and meiosis in *Rhodnius Prolixus*: The fine structure of the spindle and diffuse kinetochore. *Journal of Ultrastructure Research*, 18(5–6), 489–501. [https://doi.org/10.1016/S0022-5320\(67\)80199-0](https://doi.org/10.1016/S0022-5320(67)80199-0)
- Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology*, 109(1). <https://doi.org/10.1002/0471142727.mb2129s109>
- Carroll, C. W., Milks, K. J., & Straight, A. F. (2010). Dual recognition of CENP-A nucleosomes is required for centromere assembly. *The Journal of Cell Biology*, 189(7), 1143–1155. <https://doi.org/10.1083/jcb.201001013>
- Carroll, C. W., Silva, M. C. C., Godek, K. M., Jansen, L. E. T., & Straight, A. F. (2009). Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nature Cell Biology*, 11(7), 896–902. <https://doi.org/10.1038/ncb1899>
- Chapelle, A., Wennström, J., Hortling, H., & Ockey, C. H. (1966). ISOCHROMOSOME-X IN MAN. PART I. *Hereditas*, 54(3), 260–276. <https://doi.org/10.1111/j.1601-5223.1966.tb02021.x>
- Cheeseman, I. M. (2004). A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. *Genes & Development*, 18(18), 2255–2268. <https://doi.org/10.1101/gad.1234104>
- Chen, C.-C., Dechassa, M. L., Bettini, E., Ledoux, M. B., Belisario, C., Heun, P., Luger, K., & Mellone, B. G. (2014). CAL1 is the *Drosophila* CENP-A assembly factor. *Journal of Cell Biology*, 204(3), 313–329. <https://doi.org/10.1083/jcb.201305036>
- Chmátal, L., Gabriel, S. I., Mitsainas, G. P., Martínez-Vargas, J., Ventura, J., Searle, J. B., Schultz, R. M., & Lampson, M. A. (2014). Centromere Strength Provides the Cell Biological Basis for Meiotic Drive and Karyotype Evolution in Mice.

- Current Biology*, 24(19), 2295–2300.
<https://doi.org/10.1016/j.cub.2014.08.017>
- Cimini, D., Howell, B., Maddox, P., Khodjakov, A., Degrossi, F., & Salmon, E. D. (2001). Merotelic Kinetochore Orientation Is a Major Mechanism of Aneuploidy in Mitotic Mammalian Tissue Cells. *The Journal of Cell Biology*, 153(3), 517–528. <https://doi.org/10.1083/jcb.153.3.517>
- Clarke, L., & Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature*, 287(5782), 504–509.
- Cortes-Silva, N., Ulmer, J., Kiuchi, T., Hsieh, E., Cornilleau, G., Ladid, I., Dingli, F., Loew, D., Katsuma, S., & Drinnenberg, I. A. (2020). CenH3-Independent Kinetochore Assembly in Lepidoptera Requires CCAN, Including CENP-T. *Current Biology*, 30(4), 561-572.e10.
<https://doi.org/10.1016/j.cub.2019.12.014>
- Csankovszki, G., Collette, K., Spahl, K., Carey, J., Snyder, M., Petty, E., Patel, U., Tabuchi, T., Liu, H., McLeod, I., Thompson, J., Sarkesik, A., Yates, J., Meyer, B. J., & Hagstrom, K. (2009). Three Distinct Condensin Complexes Control *C. elegans* Chromosome Dynamics. *Current Biology*, 19(1), 9–19.
<https://doi.org/10.1016/j.cub.2008.12.006>
- Cuacos, M., H, F., F, C., & Heckman, S. (2015). Atypical centromeres in plants—What they can tell us. *Frontiers in Plant Science*, 6, 913.
- Dambacher, S., Deng, W., Hahn, M., Sadic, D., Fröhlich, J., Nuber, A., Hoischen, C., Diekmann, S., Leonhardt, H., & Schotta, G. (2012). CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. *Nucleus*, 3(1), 101–110.
<https://doi.org/10.4161/nucl.18955>
- Darlington, C. D. (1936). 121(264).

- Deal, R. B., Henikoff, J. G., & Henikoff, S. (2010). Genome-Wide Kinetics of Nucleosome Turnover Determined by Metabolic Labeling of Histones. *Science*, 328(5982), 1161–1164. <https://doi.org/10.1126/science.1186777>
- Deaton, A. M., Gómez-Rodríguez, M., Mieczkowski, J., Tolstorukov, M. Y., Kundu, S., Sadreyev, R. I., Jansen, L. E., & Kingston, R. E. (2016). Enhancer regions show high histone H3.3 turnover that changes during differentiation. *ELife*, 5, e15316. <https://doi.org/10.7554/eLife.15316>
- Dernburg, A. F. (2001). Here, There, and Everywhere: Kinetochore Function on Holocentric Chromosomes. *The Journal of Cell Biology*, 153, 6.
- Desai, A. (2003). KNL-1 directs assembly of the microtubule-binding interface of the kinetochore in *C. elegans*. *Genes & Development*, 17(19), 2421–2435. <https://doi.org/10.1101/gad.1126303>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Drinnenberg, I. A., deYoung, D., Henikoff, S., & Malik, H. S. (2014). Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *ELife*, 3, e03676. <https://doi.org/10.7554/eLife.03676>
- Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., de Montigny, J., Marck, C., Neuvéglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.-M., ... Souciet, J.-L. (2004). Genome evolution in yeasts. *Nature*, 430(6995), 35–44. <https://doi.org/10.1038/nature02579>
- Dumont, J., Oegema, K., & Desai, A. (2010). A kinetochore-independent mechanism drives anaphase chromosome separation during acentrosomal meiosis. *Nature Cell Biology*, 12(9), 894–901. <https://doi.org/10.1038/ncb2093>

- Dunleavy, E. M., Almouzni, G., & Karpen, G. H. (2011). H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G₁ phase. *Nucleus*, 2(2), 146–157. <https://doi.org/10.4161/nucl.2.2.15211>
- Dunleavy, E. M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., & Almouzni-Pettinotti, G. (2009). HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. *Cell*, 137(3), 485–497. <https://doi.org/10.1016/j.cell.2009.02.040>
- Earnshaw, W. C., & Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma*, 91(3–4), 313–321.
- Escudero, M., Márquez-Corro, J. I., & Hipp, A. L. (2016). The Phylogenetic Origins and Evolutionary History of Holocentric Chromosomes. *Systematic Botany*, 41(3), 580–585. <https://doi.org/10.1600/036364416X692442>
- Fachinetti, D., Diego Folco, H., Nechemia-Arbely, Y., Valente, L. P., Nguyen, K., Wong, A. J., Zhu, Q., Holland, A. J., Desai, A., Jansen, L. E. T., & Cleveland, D. W. (2013). A two-step mechanism for epigenetic specification of centromere identity and function. *Nature Cell Biology*, 15(9), 1056–1066. <https://doi.org/10.1038/ncb2805>
- Falk, S. J., Guo, L. Y., Sekulic, N., Smoak, E. M., Mani, T., Logsdon, G. A., Gupta, K., Jansen, L. E. T., Van Duyne, G. D., Vinogradov, S. A., Lampson, M. A., & Black, B. E. (2015). CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science*, 348(6235), 699–703. <https://doi.org/10.1126/science.1259308>
- Fitzgerald-Hayes, M., Clarke, L., & Carbon, J. (1982). Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell*, 29(1), 235–244. [https://doi.org/10.1016/0092-8674\(82\)90108-8](https://doi.org/10.1016/0092-8674(82)90108-8)
- Flemming, Walther. (1882). *Zellsubstanz, Kern und Zelltheilung*. FCW Vogel.

- Foltz, D. R., Jansen, L. E. T., Bailey, A. O., Yates, J. R., Bassett, E. A., Wood, S., Black, B. E., & Cleveland, D. W. (2009). Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell*, *137*(3), 472–484. <https://doi.org/10.1016/j.cell.2009.02.039>
- Friedman, S., & Freitag, M. (2017). Centromeres and Kinetochores. In *Centromeres and Kinetochores* (pp. 85--109). Springer.
- Fu, Y., Yang, Y., Zhang, H., Farley, G., Wang, J., Quarles, K. A., Weng, Z., & Zamore, P. D. (2018). The genome of the Hi5 germ cell line from *Trichoplusia ni*, an agricultural pest and novel model for small RNA biology. *ELife*, *7*. <https://doi.org/10.7554/eLife.31628>
- Fujita, Y., Hayashi, T., Kiyomitsu, T., Toyoda, Y., Kokubu, A., Obuse, C., & Yanagida, M. (2007). Priming of Centromere for CENP-A Recruitment by Human hMis18 α , hMis18 β , and M18BP1. *Developmental Cell*, *12*(1), 17–30. <https://doi.org/10.1016/j.devcel.2006.11.002>
- Fukagawa, T., & Earnshaw, W. C. (2014). The Centromere: Chromatin Foundation for the Kinetochores. *Developmental Cell*, *30*(5), 496–508. <https://doi.org/10.1016/j.devcel.2014.08.016>
- Furuyama, S., & Biggins, S. (2007). Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proceedings of the National Academy of Sciences*, *104*(37), 14706–14711. <https://doi.org/10.1073/pnas.0706985104>
- Gascoigne, K. E., & Cheeseman, I. M. (2013). Induced dicentric chromosome formation promotes genomic rearrangements and tumorigenesis. *Chromosome Research*, *21*(4), 407–418. <https://doi.org/10.1007/s10577-013-9368-6>
- Gassmann, R., Rechtsteiner, A., Yuen, K. W., Muroyama, A., Egelhofer, T., Gaydos, L., Barron, F., Maddox, P., Essex, A., Momen, J., Ercan, S., Lieb, J. D., Oegema, K., Strome, S., & Desai, A. (2012). An inverse relationship to

- germline transcription defines centromeric chromatin in *C. elegans*. *Nature*, 484(7395), 534–537. <https://doi.org/10.1038/nature10973>
- Goday, C., & Pimpinelli, S. (1989). Centromere organization in meiotic chromosomes of *Parascaris univalens*. *Chromosoma*, 98(3), 160–166. <https://doi.org/10.1007/BF00329679>
- Goldstein, P., & Triantaphyllou, A. C. (1980). The ultrastructure of sperm development in the plant-parasitic nematode *Meloidogyne hapla*. *Journal of Ultrastructure Research*, 71(2), 143–153. [https://doi.org/10.1016/S0022-5320\(80\)90102-1](https://doi.org/10.1016/S0022-5320(80)90102-1)
- Gong, Z., Wu, Y., Koblížková, A., Torres, G. A., Wang, K., Iovene, M., Neumann, P., Zhang, W., Novák, P., Buell, C. R., Macas, J., & Jiang, J. (2012). Repeatless and Repeat-Based Centromeres in Potato: Implications for Centromere Evolution. *The Plant Cell*, 24(9), 3559–3574. <https://doi.org/10.1105/tpc.112.100511>
- Granados, R. R., Derksen, A. C. G., & Dwyer, K. G. (1986). Replication of the *Trichoplusia ni* granulosis and nuclear polyhedrosis viruses in cell cultures. *Virology*, 152(2), 472–476. [https://doi.org/10.1016/0042-6822\(86\)90150-9](https://doi.org/10.1016/0042-6822(86)90150-9)
- Groeben, C. (2019). “... the marvelous freedom to research what one finds interesting.” *Marine Genomics*, 44, 13–23. <https://doi.org/10.1016/j.margen.2019.01.001>
- Guse, A., Carroll, C. W., Moree, B., Fuller, C. J., & Straight, A. F. (2011). In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature*, 477(7364), 354–358. <https://doi.org/10.1038/nature10379>
- Haizel, T., Lim, Y. K., Leitch, A. R., & Moore, G. (2005). Molecular analysis of holocentric centromeres of *Luzula* species. *Cytogenetic and Genome Research*, 109(1–3), 134–143. <https://doi.org/10.1159/000082392>

- Hayashi, T., Ebe, M., Nagao, K., Kokubu, A., Sajiki, K., & Yanagida, M. (2014). *Schizosaccharomyces pombe* centromere protein Mis19 links Mis16 and Mis18 to recruit CENP-A through interacting with NMD factors and the SWI/SNF complex. *Genes to Cells*, 19(7), 541–554.
<https://doi.org/10.1111/gtc.12152>
- Heckmann, S., Schroeder-Reiter, E., Kumke, K., Ma, L., Nagaki, K., Murata, M., Wanner, G., & Houben, A. (2011). Holocentric Chromosomes of *Luzula elegans* Are Characterized by a Longitudinal Centromere Groove, Chromosome Bending, and a Terminal Nucleolus Organizer Region. *Cytogenetic and Genome Research*, 134(3), 220–228.
<https://doi.org/10.1159/000327713>
- Heeger, S. (2005). Genetic interactions of separate regulatory subunits reveal the diverged *Drosophila* Cenp-C homolog. *Genes & Development*, 19(17), 2041–2053. <https://doi.org/10.1101/gad.347805>
- Henikoff, S. (2001). The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science*, 293(5532), 1098–1102.
<https://doi.org/10.1126/science.1062939>
- Heun, P., Erhardt, S., Blower, M. D., Weiss, S., Skora, A. D., & Karpen, G. H. (2006). Mislocalization of the *Drosophila* Centromere-Specific Histone CID Promotes Formation of Functional Ectopic Kinetochores. *Developmental Cell*, 10(3), 303–315. <https://doi.org/10.1016/j.devcel.2006.01.014>
- Ho, J. W. K., Jung, Y. L., Liu, T., Alver, B. H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M. Y., Appert, A., Parker, S. C. J., Gu, T., Kundaje, A., Riddle, N. C., Bishop, E., Egelhofer, T. A., Hu, S. S., Alekseyenko, A. A., Rechtsteiner, A., ... Park, P. J. (2014). Comparative analysis of metazoan chromatin organization. *Nature*, 512(7515), 449–452.
<https://doi.org/10.1038/nature13415>

- Hooff, J. J., Tromer, E., Wijk, L. M., Snel, B., & Kops, G. J. (2017). Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Reports*, *18*(9), 1559–1571. <https://doi.org/10.15252/embr.201744102>
- Hori, T., Amano, M., Suzuki, A., Backer, C. B., Welburn, J. P., Dong, Y., McEwen, B. F., Shang, W.-H., Suzuki, E., Okawa, K., Cheeseman, I. M., & Fukagawa, T. (2008). CCAN Makes Multiple Contacts with Centromeric DNA to Provide Distinct Pathways to the Outer Kinetochore. *Cell*, *135*(6), 1039–1052. <https://doi.org/10.1016/j.cell.2008.10.019>
- Howman, E. V., Fowler, K. J., Newson, A. J., Redward, S., MacDonald, A. C., Kalitsis, P., & Choo, K. H. A. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proceedings of the National Academy of Sciences*, *97*(3), 1148–1153. <https://doi.org/10.1073/pnas.97.3.1148>
- Hughes-Schrader, S., & Schrader, F. (1961). The kinetochore of the hemiptera. *Chromosoma*, *12*(1), 327–350. <https://doi.org/10.1007/BF00328928>
- Hughes-Schrader, Sally. (1944). Article: A PRIMITIVE COCCID CHROMOSOME CYCLE IN PUTO SP. *The Biological Bulletin*, *87*, 167--176.
- Iwata-Otsubo, A., Dawicki-McKenna, J. M., Akera, T., Falk, S. J., Chmátal, L., Yang, K., Sullivan, B. A., Schultz, R. M., Lampson, M. A., & Black, B. E. (2017). Expanded Satellite Repeats Amplify a Discrete CENP-A Nucleosome Assembly Site on Chromosomes that Drive in Female Meiosis. *Current Biology*, *27*(15), 2365-2373.e8. <https://doi.org/10.1016/j.cub.2017.06.069>
- Jansen, L. E. T., Black, B. E., Foltz, D. R., & Cleveland, D. W. (2007). Propagation of centromeric chromatin requires exit from mitosis. *Journal of Cell Biology*, *176*(6), 795–805. <https://doi.org/10.1083/jcb.200701066>

- Karpen, G. H., & Allshire, R. C. (1997). The case for epigenetic effects on centromere identity and function. *Trends in Genetics*, *13*(12), 489–496.
[https://doi.org/10.1016/S0168-9525\(97\)01298-5](https://doi.org/10.1016/S0168-9525(97)01298-5)
- Kasinathan, S., & Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. *Molecular Biology and Evolution*, *35*(4), 949–962.
<https://doi.org/10.1093/molbev/msy010>
- Kato, H., Jiang, J., Zhou, B.-R., Rozendaal, M., Feng, H., Ghirlando, R., Xiao, T. S., Straight, A. F., & Bai, Y. (2013). A Conserved Mechanism for Centromeric Nucleosome Recognition by Centromere Protein CENP-C. *Science*, *340*(6136), 1110–1113. <https://doi.org/10.1126/science.1235532>
- Kato, T., Sato, N., Hayama, S., Yamabuki, T., Ito, T., Miyamoto, M., Kondo, S., Nakamura, Y., & Daigo, Y. (2007). Activation of Holliday Junction Recognizing Protein Involved in the Chromosomal Stability and Immortality of Cancer Cells. *Cancer Research*, *67*(18), 8544–8553.
<https://doi.org/10.1158/0008-5472.CAN-07-1307>
- Kawamoto, M., Jouraku, A., Toyoda, A., Yokoi, K., Minakuchi, Y., Katsuma, S., Fujiyama, A., Kiuchi, T., Yamamoto, K., & Shimada, T. (2019). High-quality genome assembly of the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, *107*, 53–62. <https://doi.org/10.1016/j.ibmb.2019.02.002>
- Kobayashi, I., Tsukioka, H., Kômoto, N., Uchino, K., Sezutsu, H., Tamura, T., Kusakabe, T., & Tomita, S. (2012). SID-1 protein of *Caenorhabditis elegans* mediates uptake of dsRNA into *Bombyx* cells. *Insect Biochemistry and Molecular Biology*, *42*(2), 148–154.
<https://doi.org/10.1016/j.ibmb.2011.11.007>
- Kobayashi, N., Suzuki, Y., Schoenfeld, L. W., Müller, C. A., Nieduszynski, C., Wolfe, K. H., & Tanaka, T. U. (2015). Discovery of an Unconventional Centromere in

- Budding Yeast Redefines Evolution of Point Centromeres. *Current Biology*, 25(15), 2026–2033. <https://doi.org/10.1016/j.cub.2015.06.023>
- Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell*, 128(4), 693–705. <https://doi.org/10.1016/j.cell.2007.02.005>
- Kraushaar, D. C., Jin, W., Maunakea, A., Abraham, B., Ha, M., & Zhao, K. (2013). Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biology*, 14(10), R121. <https://doi.org/10.1186/gb-2013-14-10-r121>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- LeCain, T. J. (2019). *The Matter of History. How Things Create the Past*. Cambridge University Press.
- Lercher, M. J. (2003). Coexpression of Neighboring Genes in *Caenorhabditis Elegans* Is Mostly Due to Operons and Duplicate Genes. *Genome Research*, 13(2), 238–243. <https://doi.org/10.1101/gr.553803>
- Locke, D. P., Hillier, L. W., Warren, W. C., Worley, K. C., Nazareth, L. V., Muzny, D. M., Yang, S.-P., Wang, Z., Chinwalla, A. T., Minx, P., Mitreva, M., Cook, L., Delehaunty, K. D., Fronick, C., Schmidt, H., Fulton, L. A., Fulton, R. S., Nelson, J. O., Magrini, V., ... Wilson, R. K. (2011). Comparative and demographic analysis of orang-utan genomes. *Nature*, 469(7331), 529–533. <https://doi.org/10.1038/nature09687>
- Logsdon, G. A., Barrey, E. J., Bassett, E. A., DeNizio, J. E., Guo, L. Y., Panchenko, T., Dawicki-McKenna, J. M., Heun, P., & Black, B. E. (2015). Both tails and the centromere targeting domain of CENP-A are required for centromere establishment. *Journal of Cell Biology*, 208(5), 521–531. <https://doi.org/10.1083/jcb.201412011>

- Lu, F., Wei, Z., Luo, Y., Guo, H., Zhang, G., Xia, Q., & Wang, Y. (2019). SilkDB 3.0: Visualizing and exploring multiple levels of data for silkworm. *Nucleic Acids Research*, gkz919. <https://doi.org/10.1093/nar/gkz919>
- Ma, S., Chang, J., Wang, X., Liu, Y., Zhang, J., Lu, W., Gao, J., Shi, R., Zhao, P., & Xia, Q. (2015). CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in *Bombyx mori*. *Scientific Reports*, 4(1), 4489. <https://doi.org/10.1038/srep04489>
- Maddox, P. S., 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. *Journal of Cell Biology*, 176(6), 757–763. <https://doi.org/10.1083/jcb.200701065>
- Maddox, P. S., Oegema, K., Desai, A., & Cheeseman, I. M. (2004). “Holo”er than thou: Chromosome segregation and kinetochore function in *C. elegans*. *Chromosome Research*, 12(6), 641–653. <https://doi.org/10.1023/B:CHRO.0000036588.42225.2f>
- Malik, H. S., & Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nature Structural & Molecular Biology*, 10(11), 882–891. <https://doi.org/10.1038/nsb996>
- Malik, H. S., & Henikoff, S. (2009). Major Evolutionary Transitions in Centromere Complexity. *Cell*, 138(6), 1067–1082. <https://doi.org/10.1016/j.cell.2009.08.036>
- Mandrioli, M., & Carlo Manicardi, G. (2012). Unlocking Holocentric Chromosomes: New Perspectives from Comparative and Functional Genomics? *Current Genomics*, 13(5), 343–349. <https://doi.org/10.2174/138920212801619250>
- Marques, A., & Pedrosa-Harand, A. (2016). Holocentromere identity: From the typical mitotic linear structure to the great plasticity of meiotic holocentromeres. *Chromosoma*, 125(4), 669--681.

- Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., Pellino, M., Fuchs, J., Ma, W., Kuhlmann, M., Brandt, R., Vanzela, A. L. L., Beseda, T., Šimková, H., Pedrosa-Harand, A., & Houben, A. (2015). Holocentromeres in Rhynchospora are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(44), 13633–13638.
<https://doi.org/10.1073/pnas.1512255112>
- Masumoto, H., Nakano, M., & Ohzeki, J. (2004). *The role of CENP-B and a-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres*. 14.
- McKinley, K. L., & Cheeseman, I. M. (2016). The molecular basis for centromere identity and function. *Nature Reviews Molecular Cell Biology*, *17*(1), 16–29.
<https://doi.org/10.1038/nrm.2015.5>
- McNulty, S. M., & Sullivan, B. A. (2017). Centromere silencing mechanisms. In *Centromeres and Kinetochores* (pp. 233--255). Springer.
- Mellone, B. G., Grive, K. J., Shteyn, V., Bowers, S. R., Oderberg, I., & Karpen, G. H. (2011). Assembly of Drosophila Centromeric Chromatin Proteins during Mitosis. *PLoS Genetics*, *7*(5), e1002068.
<https://doi.org/10.1371/journal.pgen.1002068>
- Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J., Sebra, R., Peluso, P., Eid, J., Rank, D., Garcia, J., DeRisi, J. L., Smith, T., Tobias, C., Ross-Ibarra, J., Korf, I., & Chan, S. W. (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology*, *14*(1), R10.
<https://doi.org/10.1186/gb-2013-14-1-r10>
- Melters, D. P., Paliulis, L. V., Korf, I. F., & Chan, S. W. L. (2012). Holocentric chromosomes: Convergent evolution, meiotic adaptations, and genomic

- analysis. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 20(5), 579–593. <https://doi.org/10.1007/s10577-012-9292-1>
- Mendiburo, M. J., Padeken, J., Fulop, S., Schepers, A., & Heun, P. (2011). *Drosophila* CENH3 Is Sufficient for Centromere Formation. *Science*, 334(6056), 686–690. <https://doi.org/10.1126/science.1206880>
- Meraldi, P., McAinsh, A., Rheinbay, E., & Sorger, P. (2006). Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biology*, 7(3), R23. <https://doi.org/10.1186/gb-2006-7-3-r23>
- Miehe, G., Miehe, S., Kaiser, K., Jianquan, L., & Zhao, X. (2008). Status and Dynamics of the Kobresia pygmaea Ecosystem on the Tibetan Plateau. *AMBIO: A Journal of the Human Environment*, 37(4), 272–279. [https://doi.org/10.1579/0044-7447\(2008\)37\[272:SADOTK\]2.0.CO;2](https://doi.org/10.1579/0044-7447(2008)37[272:SADOTK]2.0.CO;2)
- Milks, K. J., Moree, B., & Straight, A. F. (2009). Dissection of CENP-C–directed Centromere and Kinetochore Assembly. *Molecular Biology of the Cell*, 20(19), 4246–4255. <https://doi.org/10.1091/mbc.e09-05-0378>
- Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M., Ravikumar, G., Shimomura, M., Nagamura, Y., Shin-I, T., Abe, H., Shimada, T., ... Sasaki, T. (2004). The genome sequence of silkworm, *Bombyx mori*. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 11(1), 27–35.
- Mon, H., Lee, J. M., Sato, M., & Kusakabe, T. (2017). Identification and functional analysis of outer kinetochore genes in the holocentric insect *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, 86, 1–8. <https://doi.org/10.1016/j.ibmb.2017.04.005>

- Monda, J. K., & Cheeseman, I. M. (2018). The kinetochore–microtubule interface at a glance. *Journal of Cell Science*, 131(16), jcs214577.
<https://doi.org/10.1242/jcs.214577>
- Monen, J., Maddox, P. S., Hyndman, F., Oegema, K., & Desai, A. (2005). Differential role of CENP-A in the segregation of holocentric *C. elegans* chromosomes during meiosis and mitosis. *Nature Cell Biology*, 7(12), 1248–1255.
<https://doi.org/10.1038/ncb1331>
- Moore, L. L., & Roth, M. B. (2001). Hcp-4, a Cenp-C–Like Protein in *Caenorhabditis elegans*, Is Required for Resolution of Sister Centromeres. *The Journal of Cell Biology*, 153(6), 1199–1208. <https://doi.org/10.1083/jcb.153.6.1199>
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLoS Biology*, 9(8), e1001127.
<https://doi.org/10.1371/journal.pbio.1001127>
- Moraes, I. C. R., Lermontova, I., & Schubert, I. (2011). Recognition of *A. thaliana* centromeres by heterologous CENH3 requires high similarity to the endogenous protein. *Plant Molecular Biology*, 75(3), 253–261. <https://doi.org/10.1007/s11103-010-9723-3>
- Moree, B., Meyer, C. B., Fuller, C. J., & Straight, A. F. (2011). CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. *The Journal of Cell Biology*, 194(6), 855–871.
<https://doi.org/10.1083/jcb.201106079>
- Muller, H., Gil, J., & Drinnenberg, I. A. (2019). The Impact of Centromeres on Spatial Genome Architecture. *Trends in Genetics*, 35(8), 565–578.
<https://doi.org/10.1016/j.tig.2019.05.003>
- Murakami, A., & Imai, H. T. (1974). Cytological evidence for holocentric chromosomes of the silkworms, *Bombyx mori* and *B. mandarina*,

- (Bombycidae, Lepidoptera). *Chromosoma*, 47(2), 167–178.
<https://doi.org/10.1007/BF00331804>
- Musacchio, A., & Desai, A. (2017). A Molecular View of Kinetochores Assembly and Function. *Biology*, 6(4), 5. <https://doi.org/10.3390/biology6010005>
- Nagaki, K., Kashihara, K., & Murata, M. (2005). Visualization of Diffuse Centromeres with Centromere-Specific Histone H3 in the Holocentric Plant *Luzula nivea*. *The Plant Cell*, 17(7), 1886–1893.
<https://doi.org/10.1105/tpc.105.032961>
- Nardi, I. K., Zasadzińska, E., Stellfox, M. E., Knippler, C. M., & Foltz, D. R. (2016). Licensing of Centromeric Chromatin Assembly through the Mis18 α -Mis18 β Heterotetramer. *Molecular Cell*, 61(5), 774–787.
<https://doi.org/10.1016/j.molcel.2016.02.014>
- Navarro-Mendoza, M. I., Pérez-Arques, C., Panchal, S., Nicolás, F. E., Mondo, S. J., Ganguly, P., Pangilinan, J., Grigoriev, I. V., Heitman, J., Sanyal, K., & Garre, V. (2019). Early Diverging Fungus *Mucor circinelloides* Lacks Centromeric Histone CENP-A and Displays a Mosaic of Point and Regional Centromeres. *Current Biology*, 29(22), 3791–3802.e6.
<https://doi.org/10.1016/j.cub.2019.09.024>
- Nergadze, S. G., Piras, F. M., Gamba, R., Corbo, M., Cerutti, F., McCarter, J. G. W., Cappelletti, E., Gozzo, F., Harman, R. M., Antczak, D. F., Miller, D., Scharfe, M., Pavesi, G., Raimondi, E., Sullivan, K. F., & Giulotto, E. (2018). Birth, evolution, and transmission of satellite-free mammalian centromeric domains. *Genome Research*, 28(6), 789–799. <https://doi.org/10.1101/gr.231159.117>
- Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., Novák, P., Wanner, G., & Macas, J. (2012). Stretching the Rules: Monocentric Chromosomes with Multiple Centromere Domains. *PLoS Genetics*, 8(6), e1002777. <https://doi.org/10.1371/journal.pgen.1002777>

- Nishino, T., Takeuchi, K., Gascoigne, K. E., Suzuki, A., Hori, T., Oyama, T., Morikawa, K., Cheeseman, I. M., & Fukagawa, T. (2012). CENP-T-W-S-X Forms a Unique Centromeric Chromatin Structure with a Histone-like Fold. *Cell*, *148*(3), 487–501. <https://doi.org/10.1016/j.cell.2011.11.061>
- Ockey, C. H., Wennström, J., & Chapelle, A. (1966). ISOCHROMOSOME-X IN MAN. PART II. *Hereditas*, *54*(3), 277–292. <https://doi.org/10.1111/j.1601-5223.1966.tb02022.x>
- Oliveira, L., Neumann, P., Jang, T.-S., Klemme, S., Schubert, V., Koblížková, A., Houben, A., & Macas, J. (2020). Mitotic Spindle Attachment to the Holocentric Chromosomes of *Cuscuta europaea* Does Not Correlate With the Distribution of CENH3 Chromatin. *Frontiers in Plant Science*, *10*. <https://doi.org/10.3389/fpls.2019.01799>
- Orsi, G. A., Kasinathan, S., Zentner, G. E., Henikoff, S., & Ahmad, K. (2015). Mapping Regulatory Factors by Immunoprecipitation from Native Chromatin. *Current Protocols in Molecular Biology*, *110*(1). <https://doi.org/10.1002/0471142727.mb2131s110>
- Palladino, J., Chavan, A., Sposato, A., Mason, T. D., & Mellone, B. G. (2020). Targeted De Novo Centromere Formation in *Drosophila* Reveals Plasticity and Maintenance Potential of CENP-A Chromatin. *Developmental Cell*, *52*(3), 379-394.e7. <https://doi.org/10.1016/j.devcel.2020.01.005>
- Palmer, D. K., O'Day, K., Trong, H. L., Charbonneau, H., & Margolis, R. L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(9), 3734–3738.
- Phansalkar, R., Lapierre, P., & Mellone, B. G. (2012). Evolutionary insights into the role of the essential centromere protein CAL1 in *Drosophila*. *Chromosome Research*, *20*(5), 493–504. <https://doi.org/10.1007/s10577-012-9299-7>

- Pimpinelli, S., & Goday, C. (1989). *Unusual kinetochores and chromatin diminution in Parascaris*. 5, 310--315.
- Przewlaka, M. R., Venkei, Z., Bolanos-Garcia, V. M., Debski, J., Dadlez, M., & Glover, D. M. (2011). CENP-C Is a Structural Platform for Kinetochores Assembly. *Current Biology*, 21(5), 399–405.
<https://doi.org/10.1016/j.cub.2011.02.005>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842.
<https://doi.org/10.1093/bioinformatics/btq033>
- Ramírez, F., Ryan, D. P., Grünig, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1), W160–W165. <https://doi.org/10.1093/nar/gkw257>
- Robinson, J. T. (2011). Integrative genomics viewer. *Co Rresp o n d e n Ce*, 29(1), 4.
- Ross, J. E., Woodlief, K. S., & Sullivan, B. A. (2016). Inheritance of the CENP-A chromatin domain is spatially and temporally constrained at human centromeres. *Epigenetics & Chromatin*, 9(1), 20.
<https://doi.org/10.1186/s13072-016-0071-7>
- Roure, V., Medina-Pritchard, B., Lazou, V., Rago, L., Anselm, E., Venegas, D., Jeyaprakash, A. A., & Heun, P. (2019). Reconstituting Drosophila Centromere Identity in Human Cells. *Cell Reports*, 29(2), 464-479.e5.
<https://doi.org/10.1016/j.celrep.2019.08.067>
- Sahara, K., Yoshido, A., & Yasukochi, Y. (2016). A history of chromosome identification in *Bombyx mori*. *Chromosome Science*, 7.
- Sanchez-Pulido, L., Pidoux, A. L., Ponting, C. P., & Allshire, R. C. (2009). Common Ancestry of the CENP-A Chaperones Scm3 and HJURP. *Cell*, 137(7), 1173–1174. <https://doi.org/10.1016/j.cell.2009.06.010>

- Sanyal, K., Baum, M., & Carbon, J. (2004). Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. *Proceedings of the National Academy of Sciences*, *101*(31), 11374–11379. <https://doi.org/10.1073/pnas.0404318101>
- Satzinger, Helga. (2008). Theodor and Marcella Boveri: Chromosomes and cytoplasm in heredity and development. *Nature Reviews Genetics*, *9*, 231--238.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. <https://doi.org/10.1038/nmeth.2019>
- Schrader, F. (1947). THE ROLE OF THE KINETOCHORE IN THE CHROMOSOMAL EVOLUTION OF THE HETEROPTERA AND HOMOPTERA. *Evolution*, *1*(3), 134–142. <https://doi.org/10.1111/j.1558-5646.1947.tb01332.x>
- Schrader, Franz. (1935). Notes an the mitotic behavior of long chromosomes. *Cytologia*, *6*, 422--430.
- Screpanti, E., De Antoni, A., Alushin, G. M., Petrovic, A., Melis, T., Nogales, E., & Musacchio, A. (2011). Direct Binding of Cenp-C to the Mis12 Complex Joins the Inner and Outer Kinetochores. *Current Biology*, *21*(5), 391–398. <https://doi.org/10.1016/j.cub.2010.12.039>
- Senaratne, A. P., & Drinnenberg, I. A. (2017). All that is old does not wither: Conservation of outer kinetochore proteins across all eukaryotes? *Journal of Cell Biology*, *216*(2), 291–293. <https://doi.org/10.1083/jcb.201701025>
- Senaratne, A. P., Muller, H., Fryer, K. A., & Drinnenberg, I. A. (2020). *The molecular architecture of CenH3-deficient holocentromeres in Lepidoptera is dependent*

- on transcriptional and chromatin dynamics* [Preprint]. Genomics.
<https://doi.org/10.1101/2020.07.09.193375>
- Shang, W.-H., Hori, T., Toyoda, A., Kato, J., Pependorf, K., Sakakibara, Y., Fujiyama, A., & Fukagawa, T. (2010). Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Research*, 20(9), 1219–1228. <https://doi.org/10.1101/gr.106245.110>
- Skene, P. J., & Henikoff, S. (2015). A simple method for generating high-resolution maps of genome-wide protein binding. *ELife*, 4, e09225.
<https://doi.org/10.7554/eLife.09225>
- Smith, O. K., Limouse, C., Fryer, K. A., Teran, N. A., Sundararajan, K., Heald, R., & Straight, A. F. (2020). *Identification and characterization of centromeric sequences in Xenopus laevis* [Preprint]. Genomics.
<https://doi.org/10.1101/2020.06.23.167643>
- Solis, M. A., & Pogue, M. G. (1999). Lepidopteran Biodiversity: Patterns and Estimators. *American Entomologist*, 45(4), 206–212.
<https://doi.org/10.1093/ae/45.4.206>
- Stear, J. H., & Roth, M. B. (2002). Characterization of HCP-6, a *C. elegans* protein required to prevent chromosome twisting and merotelic attachment. *Genes & Development*, 16, 1498--1508.
- Steiner, F. A., & Henikoff, S. (2014). Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *ELife*, 3, e02025. <https://doi.org/10.7554/eLife.02025>
- Steiner, F. A., & Henikoff, S. (2015). Diversity in the organization of centromeric chromatin. *Current Opinion in Genetics & Development*, 31, 28–35.
<https://doi.org/10.1016/j.gde.2015.03.010>
- Stellfox, M. E., Nardi, I. K., Knippler, C. M., & Foltz, D. R. (2016). Differential Binding Partners of the Mis18 α/β YIPPEE Domains Regulate Mis18 Complex

- Recruitment to Centromeres. *Cell Reports*, 15(10), 2127–2135. <https://doi.org/10.1016/j.celrep.2016.05.004>
- Stinchcomb, D. T., Shaw, J. E., Carr, S. H., & Hirsh, D. (1985). Extrachromosomal DNA transformation of *Caenorhabditis elegans*. *Molecular and Cellular Biology*, 5(12), 3484–3496. <https://doi.org/10.1128/MCB.5.12.3484>
- Stoler, S., Keith, K. C., Curnick, K. E., & 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 & Development*, 9(5), 573–586.
- Suetsugu, Y., Futahashi, R., Kanamori, H., Kadono-Okuda, K., Sasanuma, S., Narukawa, J., Ajimura, M., Jouraku, A., Namiki, N., Shimomura, M., Sezutsu, H., Osanai-Futahashi, M., Suzuki, M. G., Daimon, T., Shinoda, T., Taniai, K., Asaoka, K., Niwa, R., Kawaoka, S., ... Mita, K. (2013). Large Scale Full-Length cDNA Sequencing Reveals a Unique Genomic Landscape in a Lepidopteran Model Insect, *Bombyx mori*. *Genes & Genomes|Genetics*, 3(9), 1481–1492. <https://doi.org/10.1534/g3.113.006239>
- Sugimoto, K., Yata, H., Muro, Y., & Himeno, M. (1994). Human Centromere Protein C (CENP-C) Is a DNA-Binding Protein Which Possesses a Novel DNA-Binding Motif1. *The Journal of Biochemistry*, 116(4), 877–881. <https://doi.org/10.1093/oxfordjournals.jbchem.a124610>
- Sullivan, B. A., & Karpen, G. H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature Structural & Molecular Biology*, 11(11), 1076–1083. <https://doi.org/10.1038/nsmb845>
- Tachiwana, H., Müller, S., Blümer, J., Klare, K., Musacchio, A., & Almouzni, G. (2015). HJURP Involvement in De Novo CenH3CENP-A and CENP-C

- Recruitment. *Cell Reports*, 11(1), 22–32.
<https://doi.org/10.1016/j.celrep.2015.03.013>
- Talbert, P. B., Bayes, J. J., & Henikoff, S. (2009). Evolution of Centromeres and Kinetochores: A Two-Part Fugue. In P. De Wulf & W. C. Earnshaw (Eds.), *The Kinetochores*: (pp. 1–37). Springer New York. https://doi.org/10.1007/978-0-387-69076-6_7
- Talbert, P. B., Masuelli, R., Tyagi, A. P., Comai, L., & Henikoff, S. (2002). Centromeric Localization and Adaptive Evolution of an Arabidopsis Histone H3 Variant. *The Plant Cell*, 14(5), 1053–1066.
<https://doi.org/10.1105/tpc.010425>
- Talla, V., Suh, A., Kalsoom, F., Dincă, V., Vila, R., Friberg, M., Wiklund, C., & Backström, N. (2017). Rapid Increase in Genome Size as a Consequence of Transposable Element Hyperactivity in Wood-White (Leptidea) Butterflies. *Genome Biology and Evolution*, 9(10), 2491–2505.
<https://doi.org/10.1093/gbe/evx163>
- Teves, S. S., & Henikoff, S. (2014). Transcription-generated torsional stress destabilizes nucleosomes. *Nature Structural & Molecular Biology*, 21(1), 88–94. <https://doi.org/10.1038/nsmb.2723>
- Torné, J., Ray-Gallet, D., Boyarchuk, E., Garnier, M., Coulon, A., Orsi, G. A., & Almouzni, G. (2019). *Two distinct HIRA-dependent pathways handle H3.3 de novo deposition and recycling during transcription* [Preprint]. Cell Biology. <https://doi.org/10.1101/2019.12.18.880716>
- Villasante, A., Abad, J. P., & Mendez-Lago, M. (2007). Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proceedings of the National Academy of Sciences*, 104(25), 10542–10547.
<https://doi.org/10.1073/pnas.0703808104>

- Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T. L., Adelson, D. L., Bailey, E., Bellone, R. R., Blocker, H., Distl, O., Edgar, R. C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J. N., Penedo, M. C. T., Raison, J. M., ... Lindblad-Toh, K. (2009). Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science*, *326*(5954), 865–867. <https://doi.org/10.1126/science.1178158>
- Wang, J., Liu, X., Dou, Z., Chen, L., Jiang, H., Fu, C., Fu, G., Liu, D., Zhang, J., Zhu, T., Fang, J., Zang, J., Cheng, J., Teng, M., Ding, X., & Yao, X. (2014). Mitotic Regulator Mis18 β Interacts with and Specifies the Centromeric Assembly of Molecular Chaperone Holliday Junction Recognition Protein (HJURP). *Journal of Biological Chemistry*, *289*(12), 8326–8336. <https://doi.org/10.1074/jbc.M113.529958>
- Waye, J. S., & Willard, H. F. (1987). Nucleotide sequence heterogeneity of alpha satellite repetitive DNA: A survey of alphoid sequences from different human chromosomes. *Nucleic Acids Research*, *15*(18), 7549–7569. <https://doi.org/10.1093/nar/15.18.7549>
- Wei, S., HongSong, Y., YiHong, S., Yutaka, B., ZhongHuai, X., & Ze, Z. (2012). *Phylogeny and evolutionary history of the silkworm*. 14.
- Wong, C. Y. Y., Ling, Y. H., Mak, J. K. H., Zhu, J., & Yuen, K. W. Y. (2020). “Lessons from the extremes: Epigenetic and genetic regulation in point monocentromere and holocentromere establishment on artificial chromosomes.” *Experimental Cell Research*, *390*(2), 111974. <https://doi.org/10.1016/j.yexcr.2020.111974>
- Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., Pan, G., Xu, J., Liu, C., Lin, Y., Qian, J., Hou, Y., Wu, Z., Li, G., Pan, M., ... Biology Analysis Group. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science (New York, N.Y.)*, *306*(5703), 1937–1940. <https://doi.org/10.1126/science.1102210>

- Yamanaka, N., Rewitz, K. F., & O'Connor, M. B. (2013). Ecdysone Control of Developmental Transitions: Lessons from *Drosophila* Research. *Annual Review of Entomology*, 58(1), 497–516. <https://doi.org/10.1146/annurev-ento-120811-153608>
- Yuen, K. W. Y., Nabeshima, K., Oegema, K., & Desai, A. (2011). Rapid De Novo Centromere Formation Occurs Independently of Heterochromatin Protein 1 in *C. elegans* Embryos. *Current Biology*, 21(21), 1800–1807. <https://doi.org/10.1016/j.cub.2011.09.016>
- Zasadzinska, E., & Foltz, D. R. (2017). Orchestrating the specific assembly of centromeric nucleosomes. In *Centromeres and Kinetochores* (pp. 165--192). Springer.
- Zedek, F., & Bureš, P. (2018). Holocentric chromosomes: From tolerance to fragmentation to colonization of the land. *Annals of Botany*, 121(1), 9–16. <https://doi.org/10.1093/aob/mcx118>
- Zhang, J., Cong, Q., Shen, J., Opler, P. A., & Grishin, N. V. (2019). *Genomics of a complete butterfly continent* [Preprint]. Genomics. <https://doi.org/10.1101/829887>

Annex

Senaratne et al., 2020

Senaratne & Drinnenberg, 2017

The molecular architecture of CenH3-deficient holocentromeres in Lepidoptera is dependent on transcriptional and chromatin dynamics

Aruni P. Senaratne^{1,2}, H elo ise Muller^{1,2}, Kelsey A. Fryer^{3,4} and Ines A. Drinnenberg^{1,2*}

¹ Institut Curie, PSL Research University, CNRS, UMR3664, F-75005 Paris, France

² Sorbonne Universit e, Institut Curie, CNRS, UMR3664, F-75005 Paris, France

³ Department of Biochemistry, Stanford University School of Medicine, 279 Campus Drive, Beckman Center 409, Stanford, CA 94305-5307

⁴ Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305-5120

*Correspondence: ines.drinnenberg@curie.fr

Abstract

Despite their essentiality for chromosome segregation, centromeres are diverse among eukaryotes and embody two main configurations: mono- and holocentromeres, referring respectively to a localized or unrestricted distribution of centromeric activity. Previous studies revealed that holocentricity in many insects coincides with the loss of the otherwise essential centromere component CenH3 (CENP-A), suggesting a molecular link between the two events. In this study, we leveraged recently-identified centromere components to map and characterize the centromeres of *Bombyx mori*. This uncovered a robust correlation between centromere profiles and regions of low chromatin dynamics. Transcriptional perturbation experiments showed that low chromatin activity is crucial for centromere formation in *B. mori*. Our study points to a novel mechanism of centromere formation that occurs in a manner recessive to the chromosome-wide chromatin landscape. Based on similar profiles in additional Lepidoptera, we propose an evolutionarily conserved mechanism that underlies the establishment of holocentromeres through loss of centromere specificity.

Introduction

The centromere of each chromosome in a eukaryotic cell precisely defines the point of spindle fiber attachment in order to accurately partition the genome during each cell division. In the first chronological descriptions of cell division dating back to 1882, Walther Flemming observed that upon condensation, mitotic salamander chromosomes displayed a single region of narrowed chromatin to which

spindle fibers physically attached. He termed this region as the primary constriction of a chromosome, thus representing a single (mono) site of centromeric activity (Flemming, 1882). In 1888, Theodor Boveri described that the mitotic chromosomes of the roundworm, *Ascaris megacephala* behaved differently to what had been described earlier by Flemming. *Ascaris* chromosomes lacked a primary constriction and had spindle fibers attached chromosome-wide, thus representing an unrestricted distribution of centromeric activity (Boveri, 1888). These two observations are the hallmarks of the two most common centromere architectures that we know of today, which are described as monocentric and holocentric, respectively.

Most eukaryotes harbor monocentric chromosomes. Monocentromeres range from the simplest ~125 base pair point centromeres of budding yeast (Carbon and Clarke, 1984) to the regional centromeres found abundantly in mammals and plants that can span up to several megabases in size (McKinley and Cheeseman, 2016; Muller et al., 2019). Holocentric organisms are also considerably widespread and found in diverse eukaryotic lineages such as nematodes, arthropods, and flowering plants (Melters et al., 2012).

The selective forces favoring either a mono- or holocentric architecture of chromosomes are unclear. However, the nested occurrence of holocentricity within larger monocentric groups in the eukaryotic tree has supported a general consensus that holocentric chromosomes in extant organisms were derived from monocentric ancestors (Escudero et al., 2016; Melters et al., 2012). The lack of conclusive evidence pointing to reversions to the monocentric form in any eukaryotic lineage further supports the unidirectional evolution of holocentricity and suggests it to be a stable trait once adapted. To explain the evolutionary transition from mono- to holocentric chromosomes, various, often species-specific, models and evolutionary drivers have been proposed (Malik, 2002; Nagaki et al., 2005; Neumann et al., 2012). Nevertheless, the underlying mechanism remains an open question.

Studies in a few select organisms have provided details into the molecular architecture of holocentric chromosomes. Such studies relied on profiling the distribution of the widely conserved centromere-specific histone H3 variant, CenH3 (first identified as CENP-A in mammals) (Earnshaw and Rothfield, 1985; Palmer et al., 1991) and have revealed that holocentromere organization can be different on the molecular level despite their common centromeric configuration. The elaborately described holocentric model nematode *Caenorhabditis elegans* has been shown to have low occupancy CenH3 domains spanning transcriptionally-silent chromatin interspersed among discrete, high occupancy CenH3 sites (Gassmann et al., 2012; Steiner and Henikoff, 2014). In contrast, in some holocentric plants, families of satellites and/or retrotransposons (Haizel et al., 2005; Marques et al., 2015) are found to partially co-localize with centromeres, although a centromere-specific role for these repeat sequences is unknown.

The presence of holocentric chromosomes has also been reported in several lineages of insects (Melters et al., 2012; Drinnenberg et al., 2014). Here, in contrast to holocentric nematodes and plants,

protein homology predictions across mono- and holocentric insect orders revealed that multiple, independently derived holocentric insects do not possess CenH3 (Drinnenberg et al., 2014). This finding was unexpected given that the CenH3 protein is essential for chromosome segregation in most other eukaryotes. Together with its deposition factor HJURP/Scm3 (Dunleavy et al., 2009; Foltz et al., 2009) (CAL1 in flies (Chen et al., 2014)) and other kinetochore components, CenH3 takes part in a self-propagating epigenetic loop enabling the replenishment of its replication-dependent dilution at centromeres over cell divisions (Karpen and Allshire, 1997). With the exception of budding yeast and its close relatives where centromere location is defined genetically (Carbon and Clarke, 1984; Dujon et al., 2004; Kobayashi et al., 2015), the participation of CenH3 in this epigenetic loop has been demonstrated in several monocentric systems including human cells (Barnhart et al., 2011; Black et al., 2004; Carroll et al., 2010; Fachinetti et al., 2013; Guse et al., 2011; Kato et al., 2013; Logsdon et al., 2015; Tachiwana et al., 2015) and flies (Mendiburo et al., 2011; Palladino et al., 2020; Roure et al., 2019). Alternatively, a genetic basis of centromere identity has also been proposed through the recognition of tertiary DNA structures by the CenH3 incorporation machinery, with the incorporation of CenH3 itself being a by-product of this process (Kasinathan and Henikoff, 2018). Regardless of an epigenetic or genetic basis of centromere definition, the centromere-specific function of CenH3 is found to be essential in all organisms tested (Blower and Karpen, 2001; Buchwitz et al., 1999; Howman et al., 2000; Stoler et al., 1995; Talbert et al., 2002).

In addition to multiple lineages of holocentric insects, cases of CenH3 loss have also been reported in trypanosomes (Akiyoshi and Gull, 2014; Talbert et al., 2009) and in an early-diverging fungus (Hooff et al., 2017; Navarro-Mendoza et al., 2019). Interestingly, the multiple CenH3 gene loss events in several orders of insects that are notably concomitant with the occurrence of a holocentric architecture of chromosomes in each case is the first common event to be reported surrounding the loss of CenH3. This co-occurrence of holocentromeres with CenH3 loss suggests that these insects employ alternative, CenH3-independent modes of centromere specification that occur genome-wide. Given their monocentric ancestry, CenH3-deficient insects therefore represent a useful experimental tool to understand the molecular mechanism that underlies the transition to holocentric architectures in these organisms.

In this study, we aimed to determine how CenH3-deficient holocentromeres are defined in a representative insect order, Lepidoptera (butterflies and moths). We approached this question using a combination of imaging, genomics and chromatin perturbation analyses to map and characterize the centromeres in the silk moth, *Bombyx mori*. Our findings revealed a link between centromere location and regions of low chromatin activity along the entire length of *B. mori* chromosomes. Based on our results, we propose a model for how the loss of centromere specificity can lead to chromosome-wide establishment of centromeres that is non-randomly defined by underlying chromatin activity states. Our

study provides insights into the conditions governing the loss of centromere specifying components such as CenH3 over evolutionary timescales.

Results

The kinetochore forms a broad localization pattern along *B. mori* mitotic chromosomes

To understand holocentromere architecture in CenH3-deficient insects, we used a cell line derived from our representative insect model system, *B. mori*. We targeted one previously identified centromere-proximal component of the *B. mori* kinetochore, CENP-T (Cortes-Silva et al., 2020) to visualize its localization pattern on mitotic *B. mori* chromosome spreads by immunofluorescence (IF) microscopy using a custom-made antibody (Cortes-Silva et al., 2020). We observed that the CENP-T-specific IF signal formed broad localization patterns along the polar length of sister chromatids (Figure 1A), a pattern reminiscent to kinetochore staining in other holocentric organisms (Buchwitz et al., 1999). While CENP-T is an inner kinetochore component that binds directly to DNA as shown in vertebrates (Hori et al., 2008), the centromere-distal outer kinetochore components bridge centromeric DNA to the spindle apparatus, thus serving as a proxy for potential sites of spindle fiber attachment (Musacchio and Desai, 2017). To test whether the broad localization pattern we observed for CENP-T corresponds also to sites of outer kinetochore assembly, we co-stained our chromosome spreads with another custom-made antibody specific for the outer kinetochore component Dsn1 (Cortes-Silva et al., 2020). We found that the CENP-T and Dsn1 immunosignals co-localize with one another (Figure 1B), demonstrating that diffuse regions along the poleward surface of *B. mori* sister chromatids represent potential sites driving chromosome segregation during mitosis. These stainings are the first visualizations of kinetochore components along *B. mori* chromosomes, thereby building on previous cytological data (Murakami and Imai, 1974) and confirming that *B. mori* chromosomes are holocentric.

Half of the *B. mori* genome is permissive for kinetochore assembly

We proceeded to perform chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments targeting CENP-T in unsynchronized *B. mori* cell populations in order to obtain genomic resolution maps of CENP-T's distribution on chromatin. These analyses revealed broad regions of CENP-T enrichment (Figure 1C) with maximum enrichment scores of about 2-fold (log₂) over input, indicating an extensive, yet low level of CENP-T localization along chromosomes. This pattern was highly reproducible across replicates (Figure S1A and S1B). Given that the ChIP-seq libraries were prepared from asynchronous cell populations containing ~2% mitotic cells (Cortes-Silva et al., 2020) and that the *B. mori* CENP-T, as the vertebrate CENP-T homologs (Hori et al., 2008) also localizes to chromatin

during interphase (Figure S1E), CENP-T interphasic localization likely contributes to the majority of its enrichment signal.

To evaluate the extent to which CENP-T-enriched sites identified by ChIP-seq are functional kinetochore assembly sites to drive chromosome segregation during mitosis, we performed ChIP-seq to map the genome-wide distribution pattern of Dsn1. As other outer kinetochore components, Dsn1 is found on centromeric chromatin only during mitosis (Musacchio and Desai, 2017). Analogous to our co-stainings of CENP-T and Dsn1 on mitotic chromosome spreads, the ChIP-seq profile of Dsn1 also correlated well with that of CENP-T ($r = 0.8$) (Figure 1C, G).

To better characterize CENP-T-binding patterns, we used a custom domain-calling pipeline to annotate CENP-T domains. This method revealed that CENP-T domains show no obvious clustering towards the center or telomeric regions of chromosomes (Figure 1D); are of variable size (median size = 36 kb); and have a genome-wide median coverage of 54% (Figure 1E). Considering that CENP-T coverage is quantified for CENP-T ChIP-seq signal from a cell population, this number reflects the average CENP-T coverage for that population. Thus, while approximately half of the genome is CENP-T-permissive, it is possible that from one cell to the next, CENP-T occupies only a fraction of permissive sites.

Next, we analyzed the distribution of repeat sequences underlying CENP-T sites. Approximately 47% of the *B. mori* genome is comprised of repetitive DNA sequences, with around 97% of these repeats being Class I and Class II transposable elements (Kawamoto et al., 2019). To determine if CENP-T domains preferentially occur in these repetitive elements, we first intersected our annotated CENP-T domain coordinates with a database of consensus transposon sequences for *B. mori* (kind gift from the Pillai Lab, University of Geneva). This approach revealed that the proportion of transposon sequences underlying CENP-T domains is not significantly different to that in the rest of the genome (Figure 1F). Second, as a complementary approach, we searched *de novo* for any centromere-enriched repeats as described (Smith et al., 2020). Consistent with the first approach, these analyses also revealed that CENP-T-permissive sites in the *B. mori* genome are not enriched for repetitive DNA sequences (Figure S1F). Additionally, intersecting the CENP-T domain coordinates with *B. mori* gene annotations (obtained from SilkBase: <http://silkbases.ab.a.u-tokyo.ac.jp>) revealed that CENP-T domains are relatively depleted in genes (Figure 1F). Collectively, these results led us to conclude that *B. mori* centromeres are organized as broad domains that are composed of complex DNA and depleted from gene bodies.

Kinetochore attachment sites in *B. mori* are anti-correlated with actively transcribed chromatin

Given the broad distribution pattern and absence of any consensus sequence underlying CENP-T sites, it is unlikely that centromeres along *B. mori* chromosomes are defined by a specific DNA sequence.

Our use of genomics approaches to profile the underlying chromatin environment instead revealed several correlations of the distribution of CENP-T with respect to chromatin marks governing transcriptional status. We found a positive correlation between the distributions of CENP-T and tri-methylated Lysine 27 on histone H3 (H3K27me3) ($r = 0.6$) (Figure 2A, B), a histone mark that is typically associated with transcriptionally-silent heterochromatin (Kouzarides, 2007). Conversely, we found a negative correlation in the distributions of CENP-T and two histone marks that are associated with a transcriptionally-active chromatin state (Kouzarides, 2007): (i) tri-methylated Lysine 4 on histone H3 (H3K4me3) ($r = -0.2$), and (ii) tri-methylated Lysine 36 on histone H3 (H3K36me3) ($r = -0.6$) (Figure 2A, B). We also profiled the distribution of di- and tri-methylated Lysine 9 on histone H3 (H3K9me2 and H3K9me3, respectively), two histone marks that are typically associated with transcriptionally silent heterochromatin (Kouzarides, 2007). However, the resulting ChIP-seq profiles generated using multiple different antibodies and ChIP protocols for H3K9me2/3 were very similar to a histone H3 ChIP-seq profile ($r = 0.8$ for H3K9me2 vs H3; and $r = 0.7$ for H3K9me3 vs H3, respectively) (Figure S2B, C). This offered us limited confidence in the observed patterns of H3K9me2/3 in our cell line. However, as opposed to the H3K9me3-enriched heterochromatin blocks found at the pericentromeres of monocentric organisms (Sullivan and Karpen, 2004), a lack of such regions was observed in our *B. mori* cell line as evident in the absence of any chromocenters in interphase nuclei (Figure S2D).

We additionally generated mRNA-seq data for our cell line in order to assess the expression levels at *B. mori* kinetochore attachment sites. We found that mapped transcripts negatively correlated with CENP-T domains (Figure 2A). Additionally, annotated genes that did fall completely within CENP-T domains had significantly lower expression levels (median = 0.06 normalized expression units) compared to the genome-wide average (median = 1.53 normalized expression units) (Figure 2C). To account for transcripts mapping to non-annotated genes, we also looked at the expression levels across the entire genome. As with the gene-level analysis, we found that genomic regions that are enriched for CENP-T have significantly lower expression values (median = 0.2 normalized expression units) than CENP-T-depleted regions (median = 3.06 normalized expression units) (Figure 2C). Consistent with our ChIP-seq profiles for histone marks and our mRNA-seq profiles, the ChIP-seq profile of RNA Polymerase II (RNA Pol II) also revealed a negative correlation with CENP-T throughout the genome ($r = -0.64$) (Figure 2A, B). Finally, we also profiled epitope-tagged histone H3.3 driven under a constitutive promoter as a measure of nucleosome turnover (Kraushaar et al., 2013). Compared to our RNA Pol II and active histone mark profiles, we found that the H3.3 profile shows the strongest anti-correlation with CENP-T ($r = -0.76$) (Figure 2A, B). Thus, our results indicate that centromeres are excluded from genomic regions undergoing active nucleosome turnover, driven by transcription or other chromatin remodeling processes (Figure 2A, B).

Hormone-induced perturbations of gene expression result in proximal CENP-T loss or gain

To test our hypothesis that there is a causal link between *B. mori* centromere location and chromatin activity, we used an approach that allowed us to study the effects of induced transcriptional changes on CENP-T localization patterns. In insects, the ecdysone hormone response is a well-defined transcriptional response leading to the systematic activation or repression of a subset of specific genes involved in metamorphoses and development (Yamanaka et al., 2013).

We treated asynchronous *B. mori* cells with 20-Hydroxyecdysone (20E) for 48 hours and carried out mRNA-seq to determine whether our cell line had an effective transcriptional response (Figure 3A). Indeed, we found that while the expression pattern in annotated genes remained overall well-correlated before and after 20E treatment ($r = 0.99$), several genes showed differential expression upon treatment (as indicated by clusters forming away from the diagonal in the +/- 20E correlation scatterplot; Figure 3B). To identify a subset of differentially expressed genes for further analyses, we applied cut-offs in expression level to define up- and down-regulated genes (See Methods). Among a subset of twenty-six up-regulated genes, we found known 20E-responsive genes including orphan nuclear receptor genes (Yamanaka et al., 2013) (Figure 3B). We also identified one down-regulated gene that corresponded to a 20E-hydroxylase (Figure 3B). The functional annotations of these genes furthered our confidence in the specificity of the 20E response in our cell line.

Having confirmation of a visible change in expression, we next profiled CENP-T occupancy under the same conditions (Figure 3A). To identify any changes in CENP-T localization, we compared CENP-T enrichment patterns across genome-wide 10 kb windows between treated and untreated conditions. Consistent with the largely unaltered transcriptome in annotated genes (Figure 3B) and 10 kb windows (Figure 3D), global CENP-T occupancy levels were also well-correlated in treated and untreated conditions ($r = 0.9$) (Figure 3C). Nevertheless, we were able to identify specific loci with altered CENP-T levels after treatment (Figure 3C). Comparing mRNA expression levels in genomic regions showing lost or gained CENP-T binding after 20E treatment revealed that CENP-T depleted regions showed elevated expression while CENP-T enriched regions showed decreased levels of expression (Figure 3D). This is consistent with our above findings that CENP-T binding is more robust in transcriptionally silent regions.

Next, we zoomed in on each of the individual loci that lost or gained CENP-T after treatment in order to further interpret each CENP-T loss or gain event with respect to the broader chromosomal environment. We found that the most pronounced cases of CENP-T loss (representing 17 out of 48 CENP-T-depleted 10 kb genomic windows) were in fact in the proximity of two up-regulated genes. These loci corresponded to large consecutive regions of CENP-T loss after 20E treatment on chromosomes 10 and 15 where in each case, the losses were just upstream of the genes encoding for a 20E-specific nuclear receptor protein (Figure 3E). On the other hand, when we zoomed in on a genomic region on chromosome

7 that gained CENP-T after 20E treatment, we observed the opposite scenario, where the gain in CENP-T was in close proximity to a down-regulated 20E-hydroxylase gene (Figure 3E). While we cannot explain all of the observed changes in CENP-T occupancy after 20E treatment, the above described examples of CENP-T gain or loss near genes with altered expression allowed us to partially link CENP-T localization patterns to changes in transcriptional activity. The fact that altered CENP-T occupancy did not necessarily co-localize with the altered transcriptional output, but rather extended to large upstream regions, reinforces our hypothesis that changes of the chromatin landscape such as nucleosome eviction or reassembly in gene bodies, promoter or enhancer regions of up- or down-regulated genes interfere with or enable CENP-T localization, respectively.

In order to further evaluate whether it is a change in chromatin dynamics in the proximity of differential-regulated genes that underlies the changes in CENP-T occupancy, we allowed the transcriptional program to reset over the course of 5 days upon removal of the 20E hormone from the media (Figure 3A). Analyses of gene expression levels in the hormone-washout sample revealed that the previously identified subset of induced genes was mostly restored to expression levels similar to that of the control gene subset (Figure S3A). To evaluate the effects of restored gene expression on CENP-T localization, we once again profiled CENP-T occupancy 10 days following hormone-washout (Figure 3A). We found that upon 20E-washout, the prominent CENP-T losses on chromosomes 10 and 15 were recovered to levels similar to wild-type profiles (Figure 3E). The recovered events of previous CENP-T loss extended upstream of the now repressed mRNA output, once again indicating a link between chromatin activity/status and CENP-T deposition. The CENP-T gain upon 20E-treatment that was proximal to the down-regulated 20E-hydroxylase gene on chromosome 7 did not completely recover, which could be explained by incomplete restoration of expression of this gene (Figure 3E). In addition to these individual cases, similar levels of recovery following 20E-washout were observed genome-wide for regions with differential CENP-T occupancy (Figure S3B) along with comparable expression levels across all genomic windows (Figure S3C).

Taken together, our results support the hypothesis that the dynamics of the chromatin landscape determined by promoter activation or Pol II passage govern CENP-T occupancy such that CENP-T is removed from active chromatin. The complete loss of CENP-T upon promoter and gene activation also shows that no immediate recycling mechanism exists to restore CENP-T levels in those regions. Finally, the recovery of CENP-T localization upon 20E washout indicates that restoring low chromatin dynamics is sufficient for centromere formation.

Differentially expressed orthologous genes in Lepidoptera show opposite patterns of CENP-T localization

Given that the same chromosomal locus could be made permissive or repressive to CENP-T by merely changing the underlying chromatin activity status, we reasoned that such opposite patterns of CENP-T enrichment should be readily observable in an experimentally unperturbed setting, for example, on a gene that naturally varies in expression levels. To address this possibility, we turned to a close lepidopteran relative of *B. mori*, the cabbage looper *Trichoplusia ni*. We hypothesized that differentially-expressed orthologous genes between these two species provide a natural template of varying expression levels at orthologous loci to which we can correlate CENP-T occupancy levels (Figure 4A).

To first evaluate whether centromeres in *T. ni* are defined in a similar way to *B. mori*, we profiled CENP-T in an unsynchronized germ cell line derived from *T. ni* (Granados et al., 1986). We found that the distribution of *T. ni* CENP-T ChIP-seq signal resembled that of *B. mori*, wherein CENP-T localized to broad chromosomal regions (Figure S4A). Next, we did ChIP-seq profiling of the histone marks H3K27me3 and H3K36me3 in our *T. ni* cell line, which we used as markers of transcriptionally-repressed and -active chromatin, respectively. Similar to *B. mori*, in *T. ni*, the genome-wide distribution of CENP-T was positively and negatively correlated with H3K27me3 ($r = 0.5$) and H3K36me3 ($r = -0.5$), respectively (Figure S4A, B) and showed a negative correlation to mapped mRNA transcripts (Figure S4A). Thus, we concluded that the CENP-T localization in *T.ni* is correlated with silent chromatin similar to what is seen in *B. mori*.

We then reciprocally searched the *B. mori* and *T. ni* proteomes with one another to select a set of encoding orthologous genes. This revealed > 9500 orthologs to which we applied a fold-change and cut-offs in expression to select a subset of differentially-expressed genes for each species (See Methods). Accordingly, we selected 115 genes that are expressed in *B. mori* (and low in *T. ni*) and 110 genes that are expressed in *T. ni* (and low in *B. mori*) (Figure 4B). We then quantified the average CENP-T ChIP-seq scores for *B. mori* and *T. ni* over both sets of genes. In line with our previous observations, we found that CENP-T ChIP-seq signal was depleted over those genes that are expressed in *B. mori*, while CENP-T ChIP-seq signal was enriched over the corresponding lowly expressed orthologs of *T. ni* (Figure 4C). In a similar manner, we observed that CENP-T ChIP-seq signal was enriched over those genes that are lowly expressed in *B. mori*, while CENP-T ChIP-seq signal was depleted over the corresponding expressed orthologs of *T.ni*. We have thus demonstrated that across orthologous genes, CENP-T occupancy is significantly different only in those cases where there is differential expression, further supporting our hypothesis linking centromere formation to the underlying chromatin activity status.

Discussion

In this study, we characterized the CenH3-deficient holocentromere architecture of Lepidoptera, which allowed us to make an important link between their chromatin landscape and centromere distribution. Our use of a perturbation system to show that this landscape is critical for creating centromere-permissive or -dismissive environments along lepidopteran chromosomes provides insights into a new mode of centromere definition independent of CenH3.

More precisely, we identified recurrent patterns of negative correlations between the genome-wide distribution of CENP-T and factors associated with high nucleosome turnover including RNA Pol II and H3.3. We propose that lepidopteran centromere formation does not depend on an active recruitment process involving a centromere-specific epigenetic or genetic factor. Instead, kinetochores can assemble non-specifically and anywhere along the chromosomes where nucleosome turnover is low. Changes in gene expression and thus the chromatin landscape can thereafter disrupt kinetochore attachment leading to its complete loss due to the absence of an active recycling mechanism. The CENP-T profile that we thus measure corresponds to attached kinetochores that could persist over the cell cycle. This is different to the dynamics of H3.3 during transcription for instance, where a combination between new H3.3 deposition and recycling of pre-existing H3.3 coordinated by the HIRA complex enables the maintenance of the epigenetic information at that locus (Torné et al., 2019). In contrast, the presence of the lepidopteran kinetochore is not memorized and can change from cell cycle to cell cycle dependent on the chromosome-wide chromatin landscape (Figure 4D).

The organization and inheritance of the Lepidopteran holocentromere might be conceptually similar to the CenH3-encoding holocentromere in *C. elegans*. It has been proposed that actively transcribed chromosomal domains are refractory to CenH3 incorporation in *C. elegans* late-stage embryos likely due to elevated nucleosome turnover in those regions (Gassmann et al., 2012; Steiner and Henikoff, 2014). Instead, CenH3 nucleosomes can stably remain over chromosomal domains with low levels of RNA Pol II occupancy. Given the observation that *C. elegans* CenH3 was completely turned-over at each cell cycle and discontinued in the germline, the authors further proposed that CenH3 might not propagate centromere identity in this organism (Gassmann et al., 2012). The ubiquitous negative relationship between active chromatin and centromeres throughout the genomes of CenH3-deficient Lepidoptera closely resembles that of *C. elegans*. That is, based on our chromatin-perturbation experiments, we can conclude that the induction of transcriptional activity or chromatin remodeling is sufficient to induce complete dissociation of CENP-T (Figure 3). In turn, CENP-T can accumulate *de novo* in regions without requiring pre-existing CENP-T as cues (Figure 3). The potential similarities in centromere regulation in the absence or presence of CenH3 highlight the relevance of chromatin dynamics for holocentromere organization across species.

Evolutionary establishment of chromosomes with holocentric architecture in insects

The findings of this study allow us to propose a molecular mechanism underlying the mono- to holocentric transition in Lepidoptera and possibly other holocentric insects. Ancestral insects were monocentric and CenH3-dependent (Drinnenberg et al., 2014; Melters et al., 2012). CenH3 in ancestral monocentric insects might have self-propagated centromere identity at a restricted locus where it nucleated kinetochore assembly, resembling centromeres in *Drosophila melanogaster* and other monocentric organisms (Barnhart et al., 2011; Black et al., 2004; Carroll et al., 2010; Fachinetti et al., 2013; Guse et al., 2011; Karpen and Allshire, 1997; Kato et al., 2013; Logsdon et al., 2015; Mendiburo et al., 2011; Palladino et al., 2020; Roure et al., 2019; Tachiwana et al., 2015). Alternatively, centromeres could have been defined by a genetic mechanism (Kasinathan and Henikoff, 2018). In contrast, in the CenH3-deficient derived state that we characterize in this study, kinetochore assembly and thus, centromere activity occurs chromosome-wide, only antagonized by chromatin disruption processes (Figure 4D). Given that CenH3 in insects is essential in monocentric Diptera (Blower and Karpen, 2001) and its loss is only found in holocentric lineages (Drinnenberg et al., 2014), the establishment of this derived holocentric state is likely to have preceded and subsequently allowed the loss of CenH3. This is also supported by the fact that some holocentric Hemipteran insects have CenH3 homologs (Cortes-Silva et al., 2020). These Hemipterans could therefore represent an intermediate form leading to the establishment of the CenH3-deficient state.

While the molecular function of CenH3 during this possible transition state is an open question, two events must have occurred to allow the progression to a CenH3-deficient state: (i) centromere identity either conferred through an epigenetic feedback loop or through a genetic definition of centromere identity was lost, which led to the establishment of a holocentric architecture; (ii) kinetochore assembly on DNA must have become CenH3-independent, perhaps through the replacement of its ability to attach the kinetochore complex to chromatin by other kinetochore components. Future studies aiming to characterize the centromere architecture of additional holocentric insects will give more resolution to these intermediate events.

It is plausible that *C. elegans* and other holocentric nematodes that encode CenH3 may resemble this transitional stage of centromere organization. Compared to the complex CCAN of holocentric insects that mediates CenH3- and CENP-C-independent kinetochore assembly (Cortes-Silva et al., 2020), the inner kinetochores of nematodes lack CCAN but contain CenH3 and its direct binding partner CENP-C (Buchwitz et al., 1999; Cheeseman, 2004; Moore and Roth, 2001) explaining their critical roles for kinetochore attachment and outer kinetochore assembly (Carroll et al., 2010; Oegema et al., 2001; Cheeseman, 2004; Desai, 2003; Milks et al., 2009; Przewlaka et al., 2011; Screpanti et al., 2011).

Figure legends

Figure 1: Kinetochores localization patterns in *B. mori*.

A) Representative IF image of *B. mori* DAPI-stained mitotic chromosomes (blue) showing broadly-distributed immunosignal patterns of CENP-T (green). Scale bar: 5 μ m **B)** Representative IF image of *B. mori* DAPI-stained mitotic chromosomes (blue) showing the immunosignal patterns of CENP-T (red) and Dsn1 (green). Co-localization of CENP-T and Dsn1 signals can be distinguished as yellow foci in the overlay. Scale bar: 5 μ m. **C)** Genome-browser snapshot of a representative portion of *B. mori* chromosome 1 for CENP-T X-ChIP-seq, CENP-T domains, Dsn1 X-ChIP-seq and annotated genes. ChIP-seq signals are represented as histograms of the average log₂ ratio of IP/Input in genome-wide 1 kb windows. **D)** Size-scaled schematics of 28 *B. mori* chromosomes showing the distribution of CENP-T domains (dark blue segments). **E)** Features of CENP-T domains: boxplots showing the sizes (left) and the genome coverage in percent by CENP-T domains (right). **F)** Genomic features underlying CENP-T domains: barplots showing the fraction in percent of annotated interspersed repeats within CENP-T domains (dark grey) and genome-wide (light grey) (left); and the fraction in percent of annotated genes within CENP-T domains (dark grey) and genome-wide (light grey) (right). **G)** Genome-wide correlation plot of CENP-T and Dsn1 occupancy. Average log₂ ratios of IP/Input in 10 kb windows were used for plotting and calculating the pearson correlation coefficient (r) indicated on the top-left corner.

Figure 2: *B. mori* kinetochores attachment sites are anti-correlated with actively transcribed chromatin and nucleosome turnover.

A) Genome-browser snapshot of a representative portion of *B. mori* chromosome 1 for CENP-T X-ChIP-seq, CENP-T domains, H3K27me3 N-ChIP-seq, H3K4me3 X-ChIP-seq, H3K36me3 N-ChIP-seq, H3.3-3X-FLAG N-ChIP-seq, RNA Pol II X-ChIP-seq, mRNA-seq and annotated genes. ChIP-seq signals are represented as histograms of the average log₂ ratio of IP/Input in genome-wide 1 kb windows. mRNA-seq signal is represented as a histogram of log₂ normalized counts per million mapped bins (BPM). **B)** Genome-wide correlation plots comparing the occupancy of CENP-T and H3K27me3, H3K4me3, H3K36me3, H3.3-3X-FLAG and RNA Pol II. Average log₂ ratios of IP/Input in 10 kb windows were used for plotting and calculating the pearson correlation coefficient (r) indicated on the top-right corner of each plot. Comparisons between N-ChIP-seq and X-ChIP-seq or X-ChIP-seq replicates for the histone marks are shown in Figure S2A. **C)** Top: boxplot showing the expression levels in genome-wide 10 kb windows that are CENP-T-enriched (dark blue) or CENP-T-depleted (light purple), and bottom: boxplot showing the expression levels in annotated genes that are 100% within CENP-T domains (dark blue) as compared to annotated genes genome-wide (grey). Expression levels are represented as the log₂ normalized transcripts per million mapped reads (TPM). Statistical significance was tested using the

Kolmogorov-Smirnov test.

Figure 3: Hormone-induced perturbations of gene expression result in proximal CENP-T loss or gain.

A) Schematic summarizing the steps of 20E treatment and 20E washout in BmN4 cells. **B)** Genome-wide correlation plot comparing the expression levels of annotated genes (grey circles) in 20E-treated vs control (DMSO-treated) conditions. The subset of twenty-six differentially-expressed genes with user-defined expression cut-offs is highlighted (pink and blue circles, respectively). Functions of two up-regulated genes and one down-regulated gene that could be linked to three cases of differential CENP-T occupancy in the 20E-treated condition are annotated in the plot. Log₂-transformed transcripts per million mapped reads (TPM) are used to represent expression level per gene and to calculate the Pearson correlation coefficient (r) indicated on the top-left corner. **C)** Genome-wide correlation plot comparing CENP-T occupancy before (WT) and after 20E treatment in 10 kb windows (grey boxes). 10 kb windows with differential CENP-T occupancy with user defined enrichment cut-offs are demarcated (red or blue boxes, respectively). Average log₂ ratio of IP/Input in 10 kb windows were used for plotting and for calculating the Pearson correlation coefficient (r) indicated on the top-left corner. Only those 10 kb windows with total loss or total gain of CENP-T after 20E treatment from a previously enriched or depleted state, respectively were considered for our analyses. **D)** Boxplot: difference in expression levels between 20E treatment and control in genome-wide 10 kb windows (grey) and the subset of 10 kb windows with depleted or enriched CENP-T occupancy (red and blue, respectively). Difference in expression was calculated by subtracting the log₂ TPM scores of control from 20E-treated for each 10 kb window. Statistical significance was tested using the Kolmogorov-Smirnov test. **E)** Genome-browser snapshots of *B. mori* chromosomes showing two cases of CENP-T loss (left & middle) and one case of CENP-T gain (right) that could be linked to proximal changes in gene expression. Tracks are shown for CENP-T X-ChIP-seq profiles (blue) for WT, 20E-treated and 20E-washout conditions and RNA-seq profiles (orange) for control, 20E-treated and 20E-washout conditions. Pre-identified 10 kb windows with differential CENP-T occupancy (grey boxes) were found to fall consecutively within these regions. The three differentially-expressed 20E-specific genes lying upstream of these regions are highlighted with a red box in the gene track. Gene direction is marked with a black arrowhead. ChIP-seq signal is represented as the log₂ ratio of IP/Input in genome-wide 1 kb windows. RNA-seq signal is represented as log₂ normalized counts per million mapped bins (BPM). Genome-browser snapshots showing CENP-T occupancy around the remaining twenty-three pre-identified differentially-expressed genes are shown in Figure S3D.

Figure 4: Differentially expressed orthologous genes in Lepidoptera show opposite patterns of CENP-T localization.

A) Schematic summarizing the concept to identify a link between CENP-T occupancy and transcriptional activity by using naturally differentially-expressed orthologous genes of *B. mori* and *T. ni* as a template. **B)** Correlation plot showing the expression level of orthologous genes in *B. mori* and *T. ni*. Genes with similar expression levels (gray circles) and those corresponding to the subsets of the most highly expressed genes that were identified in each species which are silent in the corresponding species (purple or green circles) are indicated. Expression level per gene is represented as the log₂-transformed number of transcripts per million mapped reads (TPM) score and was used to calculate the Pearson correlation coefficient (r) indicated on the bottom-right corner. **C)** Boxplot: difference in CENP-T occupancy across all orthologs (gray) and across the subset of highly expressed genes of *B. mori* and *T. ni* (purple and green, respectively). Difference in CENP-T occupancy per gene was calculated by subtracting the average log₂ IP/Input scores for CENP-T ChIP-seq in *B. mori* from that of *T. ni*. Statistical significance was tested using the Kolmogorov-Smirnov test. **D)** Model for the architecture of CenH3-deficient holocentromeres in insects. Kinetochores bind chromosome-wide due to lack of centromere specificity. Binding is opposed only by high nucleosome turnover. Any changes to the chromatin landscape that result in alterations to the nucleosome turnover profile will also alter kinetochore attachments toward more stable regions. This kinetochore binding profile is not epigenetically inherited from one cell cycle to the next, therefore resulting in different kinetochore binding patterns for every different nucleosome turnover profile.

Acknowledgements

We thank Andreas Rechtsteiner and Nicolas Servant for helpful discussions regarding the bioinformatic pipelines used in this study, Munetaka Kawamoto and Susumu Katsuma for early access to new functional annotations for the gene models of *B. mori*, Camille Berthelot for helpful discussions regarding the analyses of orthologs, Mickaël Garnier and Patricia Le Baccon for help with imaging by microscopy, the Almouzni lab for kindly gifting us with mouse ES cells, all members of the Fachinetti and Drinnenberg labs as well members of A.P.S.'s thesis committee; Geneviève Almouzni and Leah Rosin for helpful discussions. APS receives salary support from Institut Curie and the Fondation Recherche Médicale. IAD receives salary support from the CNRS. This work is supported by the Labex DEEP ANR-11-LABX-0044 part of the IDEX Idex PSL ANR-10-IDEX-0001-02 PSL, an ATIP-AVENIR Research grant, Institut Curie and the ERC (CENEVO-758757).

References

- W. Flemming, *Zellsubstanz, Kern und Zelltheilung*, F. C. W. Vogel, 1882.
- T. Boveri, *Zellenstudien II. Die Befruchtung und Teilung des Eies von Ascaris megaloccephala* vol. 2, Gustav Fischer, 1888.
- Carbon, J., & Clarke, L. (1984). Structural and functional analysis of a yeast centromere (CEN3). *J Cell Sci*, 1984(Supplement 1), 43-58.
- Akiyoshi, B., and Gull, K. (2014). Discovery of Unconventional Kinetochores in Kinetoplastids. *Cell* 156, 1247–1258.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Barnhart, M.C., Kuich, P.H.J.L., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J. Cell Biol.* 194, 229–243.
- Black, B.E., Foltz, D.R., Chakravarthy, S., Luger, K., Woods, V.L., and Cleveland, D.W. (2004). Structural determinants for generating centromeric chromatin. *Nature* 430, 578–582.
- Blower, M.D., and Karpen, G.H. (2001). The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat. Cell Biol.* 3, 730–739.
- Buchwitz, B.J., Ahmad, K., Moore, L.L., Roth, M.B., and Henikoff, S. (1999). A histone-H3-like protein in *C. elegans*. *Nature* 401, 547–548.
- Carroll, C.W., Milks, K.J., and Straight, A.F. (2010). Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol.* 189, 1143–1155.
- Cheeseman, I.M. (2004). A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. *Genes Dev.* 18, 2255–2268.
- Chen, C.-C., Dechassa, M.L., Bettini, E., Ledoux, M.B., Belisario, C., Heun, P., Luger, K., and Mellone, B.G. (2014). CAL1 is the *Drosophila* CENP-A assembly factor. *J. Cell Biol.* 204, 313–329.
- Cortes-Silva, N., Ulmer, J., Kiuchi, T., Hsieh, E., Cornilleau, G., Ladid, I., Dingli, F., Loew, D., Katsuma, S., and Drinnenberg, I.A. (2020). CenH3-Independent Kinetochore Assembly in Lepidoptera Requires CCAN, Including CENP-T. *Curr. Biol.* 30, 561-572.e10.
- Desai, A. (2003). KNL-1 directs assembly of the microtubule-binding interface of the kinetochore in *C. elegans*. *Genes Dev.* 17, 2421–2435.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and

- Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Drinnenberg, I.A., deYoung, D., Henikoff, S., and Malik, H.S. (2014). Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *ELife* 3.
- Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., de Montigny, J., Marck, C., Neuvéglise, C., Talla, E., et al. (2004). Genome evolution in yeasts. *Nature* 430, 35–44.
- Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. *Cell* 137, 485–497.
- Earnshaw, W.C., and Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* 91, 313–321.
- Escudero, M., Márquez-Corro, J.I., and Hipp, A.L. (2016). The Phylogenetic Origins and Evolutionary History of Holocentric Chromosomes. *Syst. Bot.* 41, 580–585.
- Fachinetti, D., Diego Folco, H., Nechemia-Arbely, Y., Valente, L.P., Nguyen, K., Wong, A.J., Zhu, Q., Holland, A.J., Desai, A., Jansen, L.E.T., et al. (2013). A two-step mechanism for epigenetic specification of centromere identity and function. *Nat. Cell Biol.* 15, 1056–1066.
- Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell* 137, 472–484.
- Fu, Y., Yang, Y., Zhang, H., Farley, G., Wang, J., Quarles, K.A., Weng, Z., and Zamore, P.D. (2018). The genome of the Hi5 germ cell line from *Trichoplusia ni*, an agricultural pest and novel model for small RNA biology. *ELife* 7.
- Gassmann, R., Rechtsteiner, A., Yuen, K.W., Muroyama, A., Egelhofer, T., Gaydos, L., Barron, F., Maddox, P., Essex, A., Monen, J., et al. (2012). An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans*. *Nature* 484, 534–537.
- Granados, R.R., Derksen, A.C.G., and Dwyer, K.G. (1986). Replication of the *Trichoplusia ni* granulosis and nuclear polyhedrosis viruses in cell cultures. *Virology* 152, 472–476.
- Guse, A., Carroll, C.W., Moree, B., Fuller, C.J., and Straight, A.F. (2011). In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature* 477, 354–358.
- Haizel, T., Lim, Y.K., Leitch, A.R., and Moore, G. (2005). Molecular analysis of holocentric centromeres of *Luzula* species. *Cytogenet. Genome Res.* 109, 134–143.
- Hooff, J.J., Tromer, E., Wijk, L.M., Snel, B., and Kops, G.J. (2017). Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep.* 18, 1559–1571.
- Hori, T., Amano, M., Suzuki, A., Backer, C.B., Welburn, J.P., Dong, Y., McEwen, B.F., Shang, W.-H.,

Suzuki, E., Okawa, K., et al. (2008). CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* *135*, 1039–1052.

Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H.A. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc. Natl. Acad. Sci.* *97*, 1148–1153.

Karpen, G.H., and Allshire, R.C. (1997). The case for epigenetic effects on centromere identity and function. *Trends Genet.* *13*, 489–496.

Kasinathan, S., and Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. *Mol. Biol. Evol.* *35*, 949–962.

Kato, H., Jiang, J., Zhou, B.-R., Rozendaal, M., Feng, H., Ghirlando, R., Xiao, T.S., Straight, A.F., and Bai, Y. (2013). A Conserved Mechanism for Centromeric Nucleosome Recognition by Centromere Protein CENP-C. *Science* *340*, 1110–1113.

Kawamoto, M., Jouraku, A., Toyoda, A., Yokoi, K., Minakuchi, Y., Katsuma, S., Fujiyama, A., Kiuchi, T., Yamamoto, K., and Shimada, T. (2019). High-quality genome assembly of the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* *107*, 53–62.

Kobayashi, I., Tsukioka, H., Kômoto, N., Uchino, K., Sezutsu, H., Tamura, T., Kusakabe, T., and Tomita, S. (2012). SID-1 protein of *Caenorhabditis elegans* mediates uptake of dsRNA into *Bombyx* cells. *Insect Biochem. Mol. Biol.* *42*, 148–154.

Kobayashi, N., Suzuki, Y., Schoenfeld, L.W., Müller, C.A., Nieduszynski, C., Wolfe, K.H., and Tanaka, T.U. (2015). Discovery of an Unconventional Centromere in Budding Yeast Redefines Evolution of Point Centromeres. *Curr. Biol.* *25*, 2026–2033.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell* *128*, 693–705.

Kraushaar, D.C., Jin, W., Maunakea, A., Abraham, B., Ha, M., and Zhao, K. (2013). Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biol.* *14*, R121.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.

Logsdon, G.A., Barrey, E.J., Bassett, E.A., DeNizio, J.E., Guo, L.Y., Panchenko, T., Dawicki-McKenna, J.M., Heun, P., and Black, B.E. (2015). Both tails and the centromere targeting domain of CENP-A are required for centromere establishment. *J. Cell Biol.* *208*, 521–531.

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

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., Pellino, M., Fuchs, J., Ma, W., Kuhlmann, M., et al. (2015). Holocentromeres in *Rhynchospira* are associated with genome-wide

centromere-specific repeat arrays interspersed among euchromatin. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 13633–13638.

McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* *17*, 16–29.

Melters, D.P., Paliulis, L.V., Korf, I.F., and Chan, S.W.L. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Res.* *20*, 579–593.

Mendiburo, M.J., Padeken, J., Fulop, S., Schepers, A., and Heun, P. (2011). *Drosophila* CENH3 Is Sufficient for Centromere Formation. *Science* *334*, 686–690.

Milks, K.J., Moree, B., and Straight, A.F. (2009). Dissection of CENP-C–directed Centromere and Kinetochores Assembly. *Mol. Biol. Cell* *20*, 4246–4255.

Moore, L.L., and Roth, M.B. (2001). Hcp-4, a Cenp-C–Like Protein in *Caenorhabditis elegans*, Is Required for Resolution of Sister Centromeres. *J. Cell Biol.* *153*, 1199–1208.

Muller, H., Gil, J., and Drinnenberg, I.A. (2019). The Impact of Centromeres on Spatial Genome Architecture. *Trends Genet.* *35*, 565–578.

Murakami, A., and Imai, H.T. (1974). Cytological evidence for holocentric chromosomes of the silkworms, *Bombyx mori* and *B. mandarina*, (Bombycidae, Lepidoptera). *Chromosoma* *47*, 167–178.

Musacchio, A., and Desai, A. (2017). A Molecular View of Kinetochores Assembly and Function. *Biology* *6*.

Nagaki, K., Kashihara, K., and Murata, M. (2005). Visualization of Diffuse Centromeres with Centromere-Specific Histone H3 in the Holocentric Plant *Luzula nivea*. *Plant Cell* *17*, 1886–1893.

Navarro-Mendoza, M.I., Pérez-Arques, C., Panchal, S., Nicolás, F.E., Mondo, S.J., Ganguly, P., Pangilinan, J., Grigoriev, I.V., Heitman, J., Sanyal, K., et al. (2019). Early Diverging Fungus *Mucor circinelloides* Lacks Centromeric Histone CENP-A and Displays a Mosaic of Point and Regional Centromeres. *Curr. Biol.* *29*, 3791–3802.e6.

Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., Novák, P., Wanner, G., and Macas, J. (2012). Stretching the Rules: Monocentric Chromosomes with Multiple Centromere Domains. *PLoS Genet.* *8*, e1002777.

Oegema, K., Desai, A., Rybina, S., Kirkham, M., and Hyman, A.A. (2001). Functional Analysis of Kinetochores Assembly in *Caenorhabditis elegans*. *J. Cell Biol.* *153*, 1209–1226.

Orsi, G.A., Kasinathan, S., Zentner, G.E., Henikoff, S., and Ahmad, K. (2015). Mapping Regulatory Factors by Immunoprecipitation from Native Chromatin. *Curr. Protoc. Mol. Biol.* *110*.

Palladino, J., Chavan, A., Sposato, A., Mason, T.D., and Mellone, B.G. (2020). Targeted De Novo Centromere Formation in *Drosophila* Reveals Plasticity and Maintenance Potential of CENP-A

Chromatin. *Dev. Cell* 52, 379-394.e7.

Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., and Margolis, R.L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl. Acad. Sci. U. S. A.* 88, 3734–3738.

Przewloka, M.R., Venkei, Z., Bolanos-Garcia, V.M., Debski, J., Dadlez, M., and Glover, D.M. (2011). CENP-C Is a Structural Platform for Kinetochores Assembly. *Curr. Biol.* 21, 399–405.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165.

Robinson, J.T. (2011). Integrative genomics viewer. *C O Rresp O N N Ce* 29, 4.

Roure, V., Medina-Pritchard, B., Lazou, V., Rago, L., Anselm, E., Venegas, D., Jeyapakash, A.A., and Heun, P. (2019). Reconstituting Drosophila Centromere Identity in Human Cells. *Cell Rep.* 29, 464-479.e5.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682.

Screpanti, E., De Antoni, A., Alushin, G.M., Petrovic, A., Melis, T., Nogales, E., and Musacchio, A. (2011). Direct Binding of Cenp-C to the Mis12 Complex Joins the Inner and Outer Kinetochores. *Curr. Biol.* 21, 391–398.

Skene, P.J., and Henikoff, S. (2015). A simple method for generating high-resolution maps of genome-wide protein binding. *ELife* 4, e09225.

Smith, O.K., Limouse, C., Fryer, K.A., Teran, N.A., Sundararajan, K., Heald, R., and Straight, A.F. (2020). Identification and characterization of centromeric sequences in *Xenopus laevis* (Genomics).

Steiner, F.A., and Henikoff, S. (2014). Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *ELife* 3, e02025.

Stoler, S., Keith, K.C., Curnick, K.E., and 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, B.A., and Karpen, G.H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat. Struct. Mol. Biol.* 11, 1076–1083.

Tachiwana, H., Müller, S., Blümer, J., Klare, K., Musacchio, A., and Almouzni, G. (2015). HJURP

Involvement in De Novo CenH3CENP-A and CENP-C Recruitment. *Cell Rep.* *11*, 22–32.

Talbert, P.B., Masuelli, R., Tyagi, A.P., Comai, L., and Henikoff, S. (2002). Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. *Plant Cell* *14*, 1053–1066.

Talbert, P.B., Bayes, J.J., and Henikoff, S. (2009). Evolution of Centromeres and Kinetochores: A Two-Part Fugue. In *The Kinetochores*, P. De Wulf, and W.C. Earnshaw, eds. (New York, NY: Springer New York), pp. 1–37.

Torné, J., Ray-Gallet, D., Boyarchuk, E., Garnier, M., Coulon, A., Orsi, G.A., and Almouzni, G. (2019). Two distinct HIRA-dependent pathways handle H3.3 *de novo* deposition and recycling during transcription (*Cell Biology*).

Yamanaka, N., Rewitz, K.F., and O'Connor, M.B. (2013). Ecdysone Control of Developmental Transitions: Lessons from *Drosophila* Research. *Annu. Rev. Entomol.* *58*, 497–516.

Figure 1

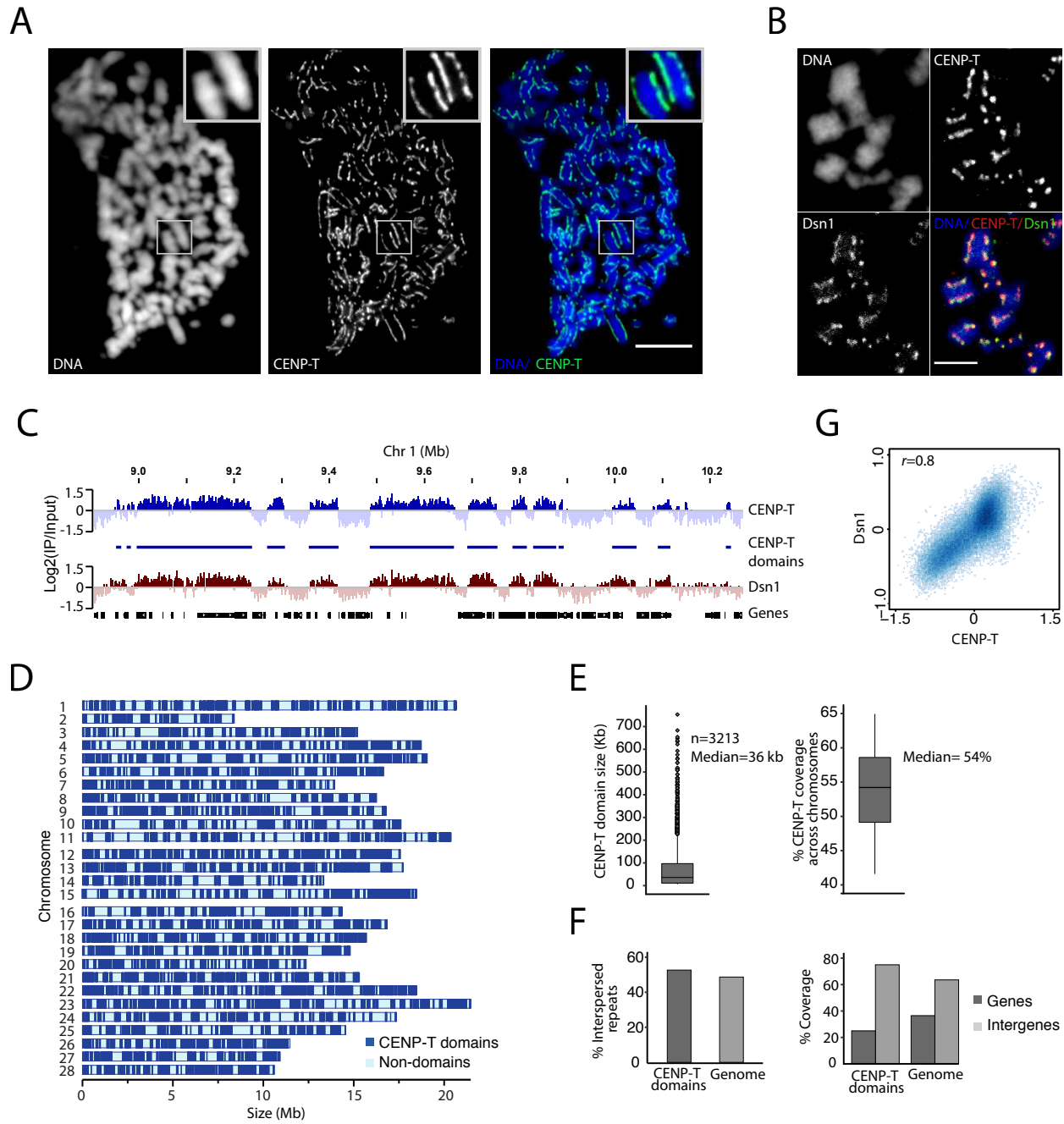


Figure 2

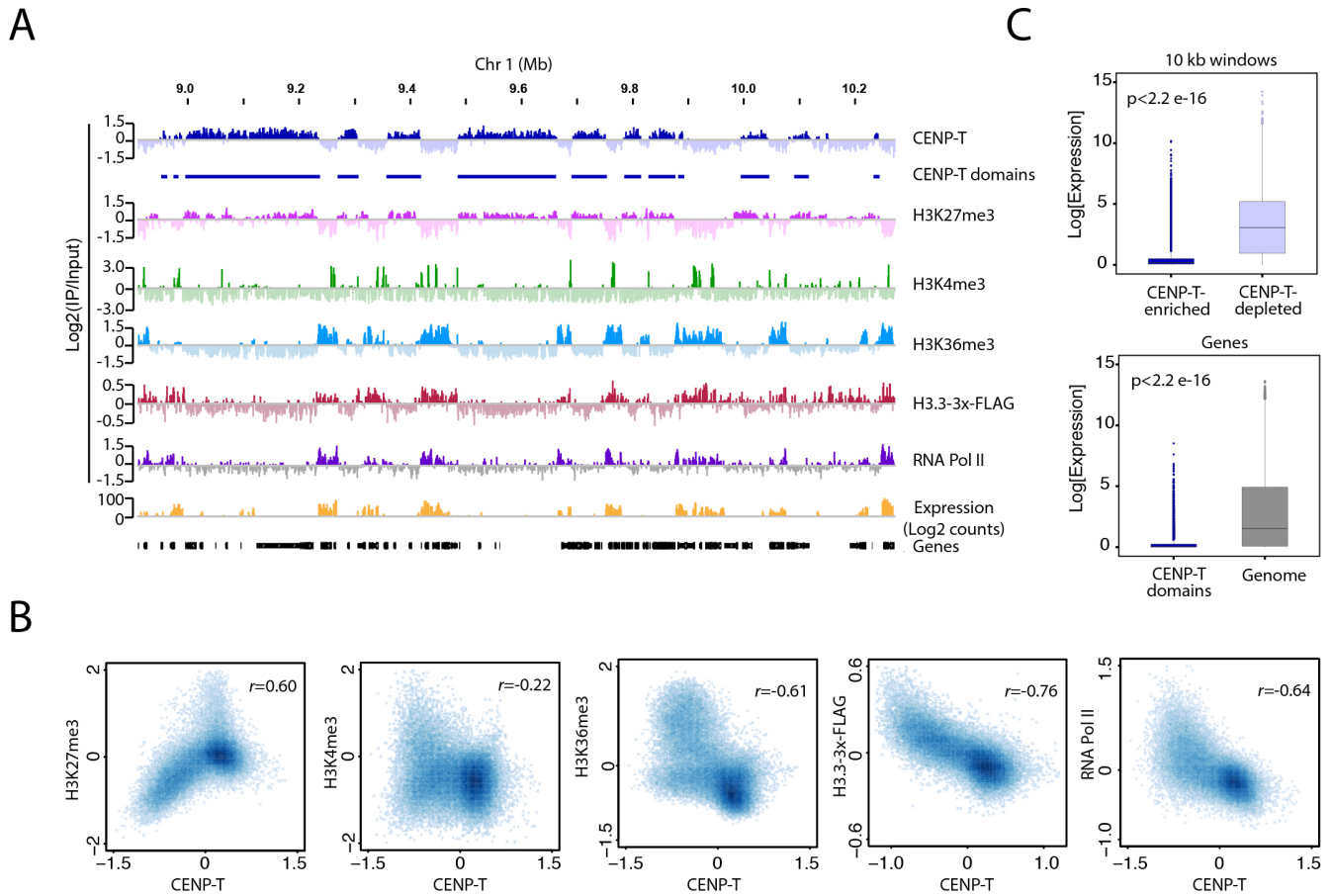
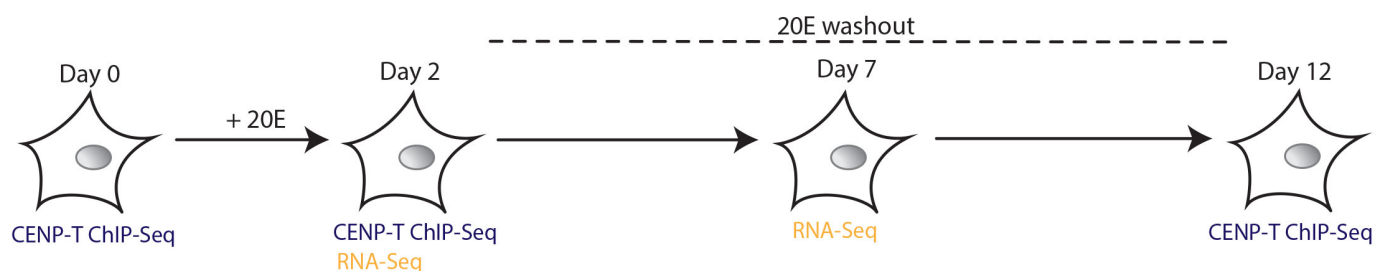
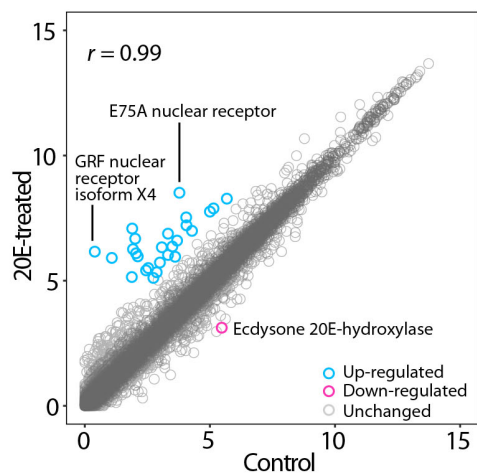


Figure 3

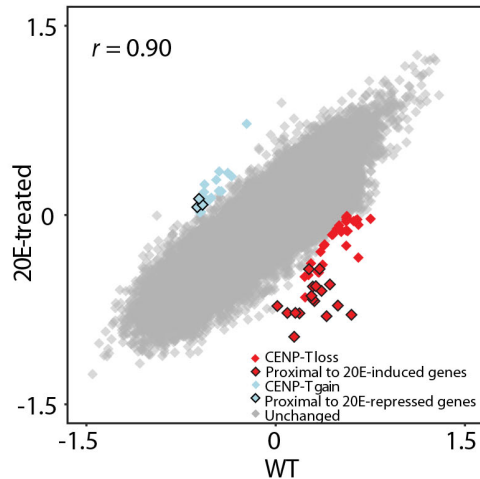
A



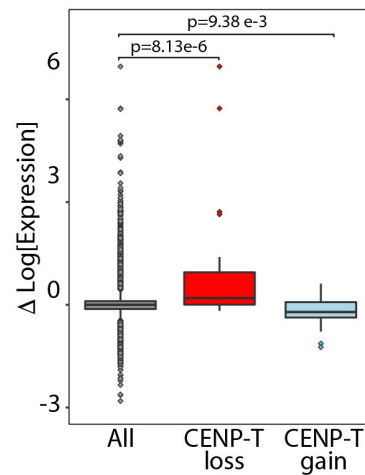
B



C



D



E

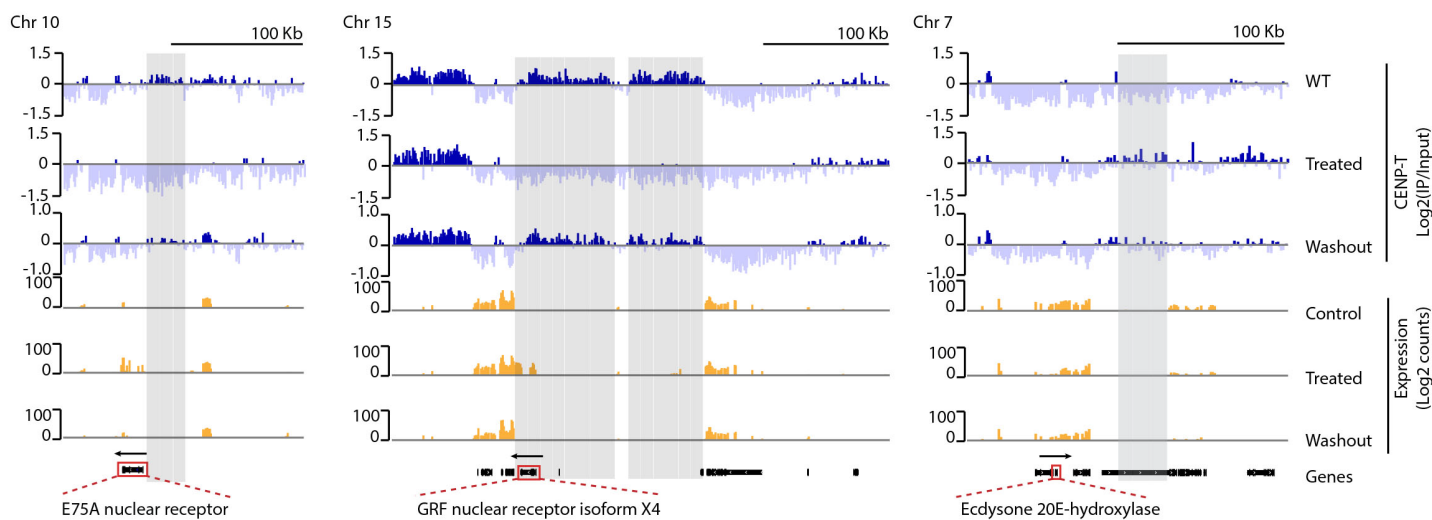
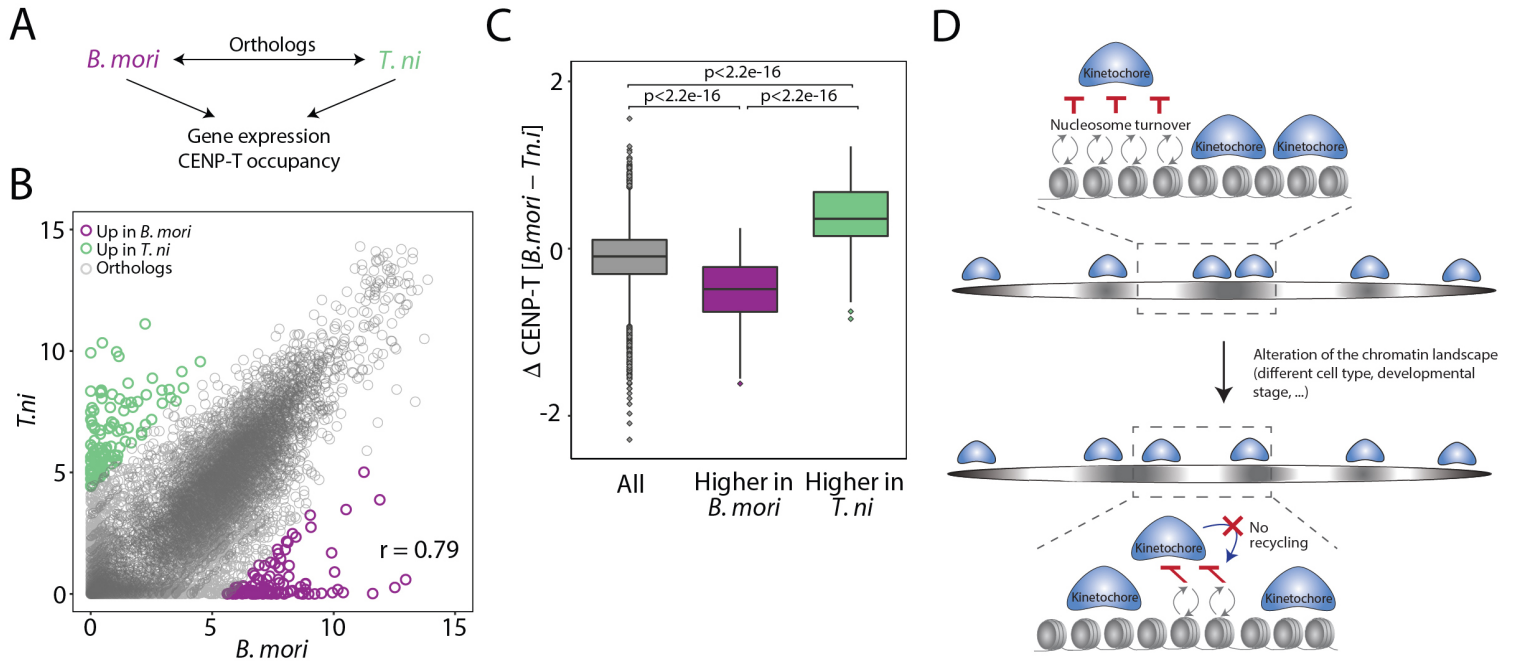


Figure 4



All that is old does not wither: Conservation of outer kinetochore proteins across all eukaryotes?

Aruni P. Senaratne^{1,2} and Ines A. Drinnenberg¹

¹Institut Curie, Paris Sciences et Lettres Research University, Centre National de la Recherche Scientifique UMR 3664, F-75005 Paris, France

²Sorbonne Universités, University Pierre-and-Marie-Curie, F-75005 Paris, France

The kinetochore drives faithful chromosome segregation in all eukaryotes, yet the underlying machinery is diverse across species. D'Archivio and Wickstead (2017. *J. Cell Biol.* <https://doi.org/10.1083/jcb.201608043>) apply sensitive homology predictions to identify proteins in kinetoplastids with similarity to canonical outer kinetochore proteins, suggesting some degree of universality in the eukaryotic kinetochore.

High-throughput sequencing technologies combined with comparative genomics have provided insights into the evolution of biological pathways. Computational prediction of homologous pathway components can trace back the ancestral origin of the underlying genes. This approach leverages knowledge of the primary amino acid sequence and is powerful if residues of the protein components are moderately conserved across species. However, its performance is limited when applied to the chromosome segregation pathway, where sequence conservation of several underlying proteins is limited. The lack of sequence conservation among some chromosome segregation components stands in stark contrast to the essentiality of this pathway. Chromosome segregation ensures the faithful transmission of genetic material from generation to generation. Crucial for this process is the kinetochore. The kinetochore is a multiprotein mosaic that assembles onto centromeric DNA to physically couple the movement of spindle microtubules to the separation of sister chromatids during anaphase.

Extensive biochemical and genetic studies in classical eukaryotic model organisms have identified a large catalog of kinetochore proteins (Cheeseman, 2014). Though similar analyses have not been performed in other organisms, computational predictions have identified homologues of several kinetochore proteins in additional species scattered across the tree of eukaryotes (Meraldi et al., 2006; Schleiffer et al., 2012). These findings reveal that most eukaryotic kinetochores consist of at least two common building blocks, namely, the histone H3 variant CenH3/CENP-A at the inner kinetochore and the Ndc80 complex at the outer kinetochore. CenH3 is enriched in centromeric chromatin at the DNA–kinetochore interface and is crucial for the initiation of kinetochore assembly (Howman et al., 2000; Blower and Karpen, 2001; Régnier et al., 2005). The Ndc80 complex binds spindle microtubules at the kinetochore–spindle interface and is crucial for driving sister chromatid separation (Kline-Smith et al., 2005). Given their widespread

conservation, it appears unexpected that computational surveys would fail to identify true homologues of canonical kinetochore proteins in kinetoplastids, a group of early-branching protozoans that include the trypanosomes.

In 2014, a pioneering study by Akiyoshi and Gull (2014) performed the first foray into the composition of kinetoplastid kinetochores. These authors applied an elegant candidate approach evaluating chromosomal localization patterns of uncharacterized proteins encoded by cell cycle–regulated genes. This led to the identification of a protein that exhibited a typical “kinetochore-like” localization behavior, termed kinetoplastid kinetochore protein 1 (KKT1). KKT1 was subsequently used as a starting point for iterative protein interaction surveys, which identified 18 additional kinetoplastid kinetochore components. While the KKT proteins are conserved among kinetoplastid species, no detectable homology to canonical kinetochore proteins could be determined, suggesting that kinetoplastids assemble their kinetochores using an alternative set of proteins. In this issue, D'Archivio and Wickstead add to this prior work and identify new kinetoplastid kinetochore proteins, one of which exhibits similarity to canonical outer kinetochore proteins.

D'Archivio and Wickstead (2017) took a reverse approach by applying remote homology predictions targeted for canonical kinetochore proteins followed by experimental validations of predicted candidates in kinetoplastids. Reasoning on a functional constraint for conservation of outer kinetochore proteins (with respect to their essential roles in forming the microtubule interface), the authors undertook a sensitive hidden Markov model (HMM)–based approach to search for remote homologues of the Ndc80 complex, Ndc80 and Nuf2. Both Ndc80 and Nuf2 have similar domain architectures consisting of an N-terminal Calponin homology (CH) fold followed by a C-terminal coiled-coil tail region (DeLuca and Musacchio, 2012). In fact, Ndc80 and Nuf2 are likely derived from a single evolutionary ancestor (Schou et al., 2013). HMM profiles constructed for the two individual protein families, separate or combined into a Ndc80/Nuf2 HMM model, were iteratively matched against proteomes of select eukaryotes. Working from true homologues into more distant evolutionary lineages, these searches identified previously undetected “Ndc80/Nuf2-like” proteins in several organisms; namely, two Excavates and the golden algae *Aureococcus anophagefferens*. Importantly, in organisms with true Ndc80/Nuf2 homologues, no additional

Correspondence to Ines A. Drinnenberg: ines.drinnenberg@curie.fr

© 2017 Senaratne and Drinnenberg This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date [see <http://www.rupress.org/terms/>]. After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).



non-homologous coiled-coil/CH fold proteins were identified, thereby indicating the specificity of the search. Next, HMM profiles containing both true Ndc80/Nuf2 homologues and newly identified hits were matched against profiles of orthologous proteins of select kinetoplastids. This search revealed additional hits with Ndc80/Nuf2-like sequence properties in these organisms. However, sequence similarity to Ndc80/Nuf2 homologues was considerably low and the contribution to detection was from alignment to coiled-coil regions of the profile. Notably, an expected CH domain was not detected in newly identified proteins.

The apparent lack of sequence similarity between canonical Ndc80/Nuf2 proteins and kinetoplastid hits meant that their role at the kinetochore–microtubule interface was still questionable and could not be inferred solely based on their computational predictions. In fact, phylogenetic analyses grouped these newly identified proteins as a separate clade distinct from all known Ndc80 and Nuf2 homologues. Acknowledging this limitation, the authors turned to experimental approaches to evaluate their candidates. As a model system, they chose *Trypanosoma brucei*, the same organism previously used by Akiyoshi and Gull (2014) for the identification and characterization of the 19 KKT proteins. Fluorescently labeling their Ndc80/Nuf2-like candidate allowed D'Archivio and Wickstead (2017) to follow its subcellular localization over the cell cycle. The authors found the localization dynamics to be very similar to KKT1, the first kinetoplastid kinetochore protein identified by Akiyoshi and Gull (2014). D'Archivio and Wickstead (2017) named their newly identified protein KKT-interacting protein 1 (KKIP1).

Further, D'Archivio and Wickstead (2017) examined the functional relevance of KKIP1 for chromosome segregation in *T. brucei*. In vertebrates and fungi, Ndc80 and Nuf2 depletion impairs kinetochore–microtubule binding, leading to aberrant chromosome partitioning and segregation defects (Kline-Smith et al., 2005). Comparably, upon KKIP1 depletion in *T. brucei*, aneuploid cells rapidly accumulated with progressing cell cycles. The authors leveraged the dispensability of *T. brucei* mini-chromosomes for cell viability to further test for chromosome loss in KKIP1-depleted cells by monitoring the maintenance of marked mini-chromosomes over cell cycles. The authors detected amplified loss rates in the range of one to two orders of magnitude. Overall, their observations are similar to those seen for Ndc80- and Nuf2-compromised cells in other organisms (Kline-Smith et al., 2005). However, D'Archivio and Wickstead (2017) found impaired spindle assembly in KKIP1-depleted *T. brucei* cells—a defect not observed in other organisms. While the mechanistic link is unclear, the authors hypothesize that, in *T. brucei*, spindles are perhaps unstable when not associated with kinetochores.

D'Archivio and Wickstead (2017) next addressed the functional relationship of KKIP1 to the KKT proteins (Akiyoshi and Gull, 2014). The authors performed semiquantitative cross-linking affinity purifications under native, low, and high formaldehyde conditions and mass spectrometry to identify KKIP1 interacting partners. This approach revealed a significant enrichment of several KKT proteins as well as a nuclear pore complex component known to associate with spindles during mitosis. The central mitotic kinase, Aurora B, was also identified, further supporting participation of KKIP1 in the chromosome segregation machinery. Interestingly, the centromere-proximal proteins KKT2 and KKT3, as well as KKT13 that reaches peak levels during S phase (Akiyoshi and

Gull, 2014), were not among the potential interaction partners. Collectively with the protein localization studies, these results support a centromere-distal localization of KKIP1, enriched predominantly during mitosis.

In addition, the proteomic analyses identified a new set of potential kinetochore proteins in *T. brucei*. D'Archivio and Wickstead (2017) used the same approaches to characterize the localization and function of these proteins as they did for KKIP1, which allowed them to narrow the list down to six potential interactors, named KKIP2 to 7. While none of these proteins showed any recognizable homologues in species outside the kinetoplastids, KKIP7 was predicted to contain a phosphatase domain belonging to the family that includes members of other known mitotic phosphatases. Correct kinetochore assembly and spindle attachment in other eukaryotes are regulated by the interplay of mitotic kinases and phosphatases that modify proteins of the kinetochore (Reinhardt and Yaffe, 2013). D'Archivio and Wickstead (2017) speculate that KKIP7 acts as an antagonist of trypanosomal mitotic kinases (Aurora B and KKT kinases) to regulate phosphorylation-dependent kinetochore function.

Additional parallels to the outer kinetochore complex could be drawn from insights into the kinetoplastid kinetochore assembly cascade. In other eukaryotes, kinetochore assembly happens in an ordered manner, with the assembly of inner components preceding that of outer ones (Cheeseman, 2014). Consistent with an analogous, centromere-distal arrangement of KKIP1 (and KKIP4), D'Archivio and Wickstead (2017) found that the localization of KKIP1 occurs downstream of most representative KKT members. Conversely, other KKIP proteins were found to be dependent on KKIP1 for recruitment, indicating the upstream localization of KKIP1 in the assembly hierarchy of centromere-distal proteins in kinetoplastids.

To directly observe the arrangement of KKIP1 relative to centromere-proximal proteins such as KKT2, D'Archivio and Wickstead (2017) applied two-color fluorescence microscopy on relaxed kinetochores in anaphase cells. Consistent with its localization closer to the centromere, KKT2 appeared to be significantly skewed away from the spindle pole compared to KKIP1. The calculated distance between the two proteins was similar to the estimated thickness of kinetochore-like plaques observed by electron microscopy in Trypanosomes (Ogbadoyi et al., 2000). Thus, this arrangement of centromere-proximal and -distal proteins recapitulates the size of the *T. brucei* kinetochore complex.

This study by D'Archivio and Wickstead (2017) gives new insights into conserved principles of kinetochore composition and structure. Proteins of the Ndc80 complex are among the most conserved kinetochore components across eukaryotes, yet homologues have not been identified in kinetoplastids. Using bioinformatics analyses and experimental validations, the authors identified a novel kinetoplastid outer kinetochore component with some structural and functional similarity to Ndc80/Nuf2 homologues. Still, it is challenging to provide evidence of homology for several reasons. First, at primary amino acid sequence level, a high degree of sequence divergence is observed. Second, at structural and functional levels, the essential microtubule-binding interfaces found in canonical Ndc80—the N-terminal tail domain, the CH domain, and the characteristic microtubule binding loop region following the CH domain (Varma and Salmon, 2012)—are not detected in the kinetoplastid Ndc80/Nuf2-like candidate, KKIP1. To this end, it is still unclear how kinetoplastid outer kinetochore proteins make

essential microtubule contacts in the absence of otherwise indispensable functional motifs. As far as a universal chromosome segregation model is concerned, the findings from D'Archivio and Wickstead (2017) show that outer kinetochore proteins with recurring structural motifs such as coiled-coil domains are constitutive members of eukaryotic kinetochores (Westermann and Schleiffer, 2013). This thereby indicates some degree of universality of the eukaryotic outer kinetochore complex, particularly with regard to the presence of conserved secondary structures. In addition, this study also proves the potential of D'Archivio and Wickstead's approach for characterizing kinetochore proteins in divergent eukaryotic lineages, which may not have been detected with classical homology searches.

Acknowledgments

We thank Alberto Gatto and all the members of the Drinnenberg laboratory for helpful suggestions.

A.P. Senaratne was funded by the IC-3i International Fellowship Program. I.A. Drinnenberg was funded by an ATIP-AVENIR research grant, Institut Curie, and the Centre National de la Recherche Scientifique.

The authors declare no competing financial interests.

References

Akiyoshi, B., and K. Gull. 2014. Discovery of unconventional kinetochores in kinetoplastids. *Cell*. 156:1247–1258. <http://dx.doi.org/10.1016/j.cell.2014.01.049>

Blower, M.D., and G.H. Karpen. 2001. The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat. Cell Biol.* 3:730–739. <http://dx.doi.org/10.1038/35087045>

Cheeseman, I.M. 2014. The kinetochore. *Cold Spring Harb. Perspect. Biol.* 6:a015826. <http://dx.doi.org/10.1101/cshperspect.a015826>

D'Archivio, S., and B. Wickstead. 2017. Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J. Cell Biol.* <http://dx.doi.org/10.1083/jcb.201608043>

DeLuca, J.G., and A. Musacchio. 2012. Structural organization of the kinetochore–microtubule interface. *Curr. Opin. Cell Biol.* 24:48–56. <http://dx.doi.org/10.1016/j.ceb.2011.11.003>

Howman, E.V., K.J. Fowler, A.J. Newson, S. Redward, A.C. MacDonald, P. Kalitsis, and K.H. Choo. 2000. Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc. Natl. Acad. Sci. USA.* 97:1148–1153. <http://dx.doi.org/10.1073/pnas.97.3.1148>

Kline-Smith, S.L., S. Sandall, and A. Desai. 2005. Kinetochore–spindle microtubule interactions during mitosis. *Curr. Opin. Cell Biol.* 17:35–46. <http://dx.doi.org/10.1016/j.ceb.2004.12.009>

Meraldi, P., A.D. McAinsh, E. Rheinbay, and P.K. Sorger. 2006. Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol.* 7:R23. <http://dx.doi.org/10.1186/gb-2006-7-3-r23>

Ogbadoyi, E., K. Ersfeld, D. Robinson, T. Sherwin, and K. Gull. 2000. Architecture of the *Trypanosoma brucei* nucleus during interphase and mitosis. *Chromosoma.* 108:501–513. <http://dx.doi.org/10.1007/s004120050402>

Régnier, V., P. Vagnarelli, T. Fukagawa, T. Zerjal, E. Burns, D. Trouche, W. Earnshaw, and W. Brown. 2005. CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. *Mol. Cell. Biol.* 25:3967–3981. <http://dx.doi.org/10.1128/MCB.25.10.3967-3981.2005>

Reinhardt, H.C., and M.B. Yaffe. 2013. Phospho-Ser/Thr-binding domains: Navigating the cell cycle and DNA damage response. *Nat. Rev. Mol. Cell Biol.* 14:563–580. <http://dx.doi.org/10.1038/nrm3640>

Schleiffer, A., M. Maier, G. Litos, F. Lampert, P. Hornung, K. Mechtler, and S. Westermann. 2012. CENP-T proteins are conserved centromere receptors of the Ndc80 complex. *Nat. Cell Biol.* 14:604–613. <http://dx.doi.org/10.1038/ncb2493>

Schou, K.B., J.S. Andersen, and L.B. Pedersen. 2013. A divergent calponin homology (NN-CH) domain defines a novel family: implications for evolution of ciliary IFT complex B proteins. *Bioinformatics.* 30:899–902.

Varma, D., and E.D. Salmon. 2012. The KMN protein network—chief conductors of the kinetochore orchestra. *J. Cell Sci.* 125:5927–5936. <http://dx.doi.org/10.1242/jcs.093724>

Westermann, S., and A. Schleiffer. 2013. Family matters: Structural and functional conservation of centromere-associated proteins from yeast to humans. *Trends Cell Biol.* 23:260–269. <http://dx.doi.org/10.1016/j.tcb.2013.01.010>