The impact of centromeres on spatial genome architecture
Héloïse Muller, Ines Anna Drinnenberg, José Gil Jr.

To cite this version:
Héloïse Muller, Ines Anna Drinnenberg, José Gil Jr.. The impact of centromeres on spatial genome architecture. Trends in Genetics, Elsevier, 2019, 35 (8), pp.565-578. 10.1016/j.tig.2019.05.003. hal-02189346

HAL Id: hal-02189346
https://hal.archives-ouvertes.fr/hal-02189346
Submitted on 19 Jul 2019

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.
The impact of centromeres on spatial genome architecture

Héloïse Muller\textsuperscript{1,2*}, José Gil Jr.\textsuperscript{1,2} and Ines Anna Drinnenberg\textsuperscript{1,2}

\textsuperscript{1} Institut Curie, PSL Research University, CNRS, UMR3664, F-75005 Paris, France
\textsuperscript{2} Sorbonne Université, Institut Curie, CNRS, UMR3664, F-75005 Paris, France

\*Corresponding author: Muller, H. (heloise.muller@curie.fr)

Keywords: Centromere, chromosome architecture, nuclear compartments, microscopy, Hi-C

Abstract: The development of new technologies and experimental techniques are enabling researchers to see what was once unable to be seen. For example, the centromere was first seen as the mediator between spindle fiber and chromosome during mitosis and meiosis. Although this continues to be its most prominent role, we now know that the centromere functions beyond cellular division with important roles in genome organization and transcriptional regulation. Here we aim to share the structures and functions of centromeres in different organisms beginning with the diversity of their DNA sequence anatomies. We zoom out to describe their position in the nucleus and ultimately detail the different ways they contribute to genome organization and regulation at the spatial level.
Centromeres: Beyond chromosome segregation

Eukaryotic genomes are not randomly organized. In fact, chromatin is organized into multiple domains juxtaposed one after the other along the length of a chromosome and positioned into particular layouts and within particular regions of the nuclear space [1,2]. Among these chromatin domains, centromeres are universal to all eukaryotic chromosomes. First cytologically described in 1882 by Walther Flemming as the primary constrictions on condensed chromosomes attaching to spindle fibers [3], centromeres are essential for cell division by ensuring the equal partitioning of DNA into daughter cells.

Although their crucial function comes into play during mitosis and meiosis, centromeres are not inert during interphase. In fact, their importance in the genomes’ 3D organization and regulation has only started to be recognized. In diverse eukaryotes, centromeres occupy particular domains in interphase nuclei. This non-random organization of centromeres in turn impacts chromatin-based processes such as transcription and replication. Here, we review centromere composition, localization, and interaction patterns across different eukaryotes. Simultaneously, we highlight conserved and divergent principles of linear and spatial centromeric architectures and their relationship with genome organization and function. For this, we survey observations from microscopy and more recent genome conformation studies to provide a systematic overview on spatial centromere organization in diverse eukaryotes. We also discuss the functional relevance of particular centromeric arrangements in interphase chromatin and emphasize current limitations and alternate strategies of genomic approaches in determining the spatial organization of centromeres.

Centromeres are enigmas of many genome assemblies

Despite their essential function in mediating and controlling chromosome segregation, centromere architecture is remarkably diverse among different organisms. In fact, the length of centromeres ranges from as small as 120 base pairs to up to several megabases of DNA [4]. With some recent exceptions, sequence information is only available for shorter centromeres whereas assemblies of larger centromeres are limited due to their length and repetitive sequence composition (Figure 1, Box1).
Centromere sequences of many yeasts and fungal species have been characterized (Figure 1, Box1). Comparative genomic analyses show that centromere size and synteny (see glossary) are mostly conserved within both the Saccharomycetaceae and Candida clades, pointing toward a common origin in each of the two clades. Nevertheless, centromere sequences have rapidly diverged with very little or no detectable homology found between species [5–10]. In contrast to the budding yeasts, the well-studied fission yeast Schizosaccharomyces pombe has an incomplete genome assembly for two out of its three centromeres with the number of pericentromeric inverted repeats still unknown (Figure 1, Box 1) (www.pombase.org). Compared to fungi, sequence assemblies of the megabase-long animal and plant centromeres, composed of satellite DNA interspersed with transposable elements, are even more limited and remain largely unresolved even after using long-read sequencing approaches [11–13]. Still, in some plant organisms, including Oryzae sativa and Arabidopsis thaliana, the molecular structure of large centromeric regions for several chromosomes has been determined (Figure 1, Box 1) [14–18]. In addition, a recent study characterized the organization and sequence composition of all native Drosophila centromeres, confirming the structure unraveled by original mapping efforts on the Dp1230 minichromosome (Figure 1, Box 1) [19,20]. Finally, the complete sequence of the first human centromere has also been released by assembling the higher-order-repeat (HOR) structure of the human chromosome Y satellite array (Figure 1, Box 1) [21].

Note that some species, including chicken, horse and potato, have repeat-based centromeres as well as repeat-free centromeres depending on the chromosome [22–24]. For instance, while most centromeres from the chicken macrochromosomes are composed of chromosome-specific homogenous arrays of tandem repeats, chromosomes 5, 27 and Z are composed of unique sequences. Comparative genomic studies show that the latter are likely to represent evolutionarily more recent and “immature” centromeres that could represent an intermediate state [25]. Consistent with this, neocentromeres (see glossary) often form on sequences devoid of repetitive DNA [25] and tandem repeats have only recently been accumulating in some evolutionary new centromeres (see glossary). Regardless of their evolutionary framework, the non-repetitive nature of these centromeres provides a unique opportunity allowing their unambiguous assembly.
Seeing is believing: Centromeres follow a distinct nuclear organization

While centromere assemblies are often limited, the location of centromeres can be visualized by the presence of the histone variant CenH3 (also called CENP-A in humans), which almost universally marks centromeric chromatin in eukaryotes [26]. Microscopy studies provided the first insights into centromere organization in interphase nuclei. In numerous fungal species, including S. cerevisiae and S. pombe, centromeres cluster all along the cell cycle, which are visible as one large focus in the vicinity of the Spindle Pole Body (SPB) and opposite to the nucleolus (Figure 2) [27–30]. Nevertheless, this organization can also vary in some fungi. In C. neafomans, centromeres do not cluster in premeiotic cells and are instead separately positioned adjacent to the nuclear envelope (Figure 2). Only when cells progress toward mitosis do the centromeres gradually coalesce to a single cluster [31].

Chromosome organization in most plants has classically been divided into two main categories: Rabl or non-Rabl. The Rabl organization is characterized by 1) all centromeres being restricted to one pole of the nucleus, opposite to the nucleolus and 2) the telomeres being more dispersed in the opposite pole. This organization is observed in wheat, rye, barley and oat root tip cells (Figure 2) [32]. In contrast, sorghum and rice root cells adopt a non-Rabl conformation with centromeres and telomeres dispersed and intermingled in the nuclear area (Figure 2) [32]. Some species like maize have an intermediate conformation between Rabl and non-Rabl in nuclei from root tip cells [32]. In addition, later studies modulated this original classification by describing that chromosome organization can vary between different cell types. For example, most root nuclei of rice show non-Rabl organization, as described above, while in the xylem vessel cells centromeres and telomeres localize to opposite poles of the nuclei, consistent with a Rabl configuration [32,33].

Compared to Rabl and non-Rabl conformation, A. thaliana has a distinct centromere organization with (peri-)centromeric heterochromatin organized as clearly distinguishable chromocenters (see glossary) in the nucleus (Figure 2). These are not randomly positioned inside the nucleus and are found near the periphery or close to the nucleolus. Moreover, the number of chromocenters observed is often (in ~88% of observed nuclei) less than 10 (corresponding to 2n = 10 chromosomes), demonstrating a propensity to associate together [34].
Centromere clustering in interphase nuclei has also been observed in several animal model organisms. Counting the number of CENH3$^{CID}$ positive foci in interphase nuclei using immunofluorescence experiments demonstrated that centromeres cluster in fly cells [35]. In particular, S2 aneuploid cells (13 stable chromosomes) show 4-6 clusters (Figure 2) and third-instar larval diploid (2n=8) hemocytes show 2-3 clusters, which are localized close to the periphery of the nucleolus in both cell types.

Centromere clustering is also a common feature in human and mouse. Early studies on mouse B and T lymphocytes describe clustering of centromeres and their colocalization with chromocenters comprised of pericentromeric heterochromatin (Figure 2) [36,37]. More recent analyses of non-dividing/quiescent (G0) lymphocyte cells from mouse and humans reveal that centromeres cluster preferentially at the nuclear periphery, forming on average 13 or 9 signals per human (2n=46) or mouse (2n=40) nuclei, respectively [38]. In the few cases of a central localization, clusters are found to be associated with the nucleolus. Refining centromere locations within the nuclear region occupied by their corresponding chromosome (also called chromosome territory (CT, see glossary), centromeres are found at the periphery of CTs in 95% of studied cases and positioned towards the nuclear periphery. Other studies confirm and expand these results showing that centromere clustering is a common feature observed to various extents in all cell types analyzed [39–42], but changes dynamically over the course of the cell cycle [41,42]. The cell types analyzed include hematopoietic progenitors and mature blood cells (B and T lymphocytes, granulocytes, monocytes) as models for cell differentiation, as well as fibroblasts, which offer a comparison with a terminally differentiated cell from another lineage. Upon exit of mitosis, in early G1, very little to no clustering is observed and the number of kinetochore signals decrease in late G1 and S phases indicating centromere clustering. Over the course of G2, centromeres separate out again as chromosomes start to condense in preparation for mitosis. In addition, peripheral localization and clustering of centromeres are less pronounced in nuclei of cycling cells or progenitor cells compared to non-cycling and terminally differentiated cells [41,42]. In particular, while centromere clustering is maintained in human embryonic stem cells, their peripheral localization is mainly lost, possibly due to rapid divisions of these cells [43].
In summary, while the insights gained from microscopy approaches are limited, these studies reveal that centromeres are often confined to or even clustered in specific nuclear locations in many organisms. Their organization within the nucleus, however, varies from one species to another.

Centromere interactions drive genome 3D organization and regulation

Genome-wide chromosome architecture studies using high-throughput molecular biology techniques, like Chromosome Conformation Capture (3C, see glossary) and its derivatives (4C, Hi-C,...; see glossary) offer new ways to study the importance of centromeres in genome organization. Early studies define the top of genome organization hierarchy as being CTs [44]. In line with that, Hi-C data from different organisms describe that intrachromosomal interactions are always more frequent than interchromosomal interactions. In many organisms, and especially in mammals, chromosomes can be further segregated into two compartments: the A compartment, which includes regions of the genome enriched in genes and histone modifications associated with active chromatin, or the B compartment, which is enriched in histone modifications associated with inactive chromatin [45]. Going down the chromosome organization to the scale of several hundred kilobases, one can find domains which interact more frequently within themselves than with neighboring regions, often referred to as topologically associated domains (TADs) [46]. In addition, other signatures of genome organization are also visible on Hi-C maps, and account for specific and strong landmarks of 3D genome organization. Here, the centromere is a classic example of a region with strong topological impact. Assembling data from many organisms published to date, we will discuss at least two implications to previously described centromere interactions visible on Hi-C matrices: (i) centromeres build a specific sub-compartment inside the nucleus and (ii) they form a barrier to intrachromosomal arm interactions.

Centromeres form strong interacting sub-compartments

Three-dimensional genome organization has been studied in several fungal species. The first S. cerevisiae full genome contact map (see glossary) was published using a high-throughput 4C-derived experiment [47]. From this map, centromere clustering was reported to be the most striking feature of interchromosomal contacts, in accordance with the known
Rab1-like organization of budding yeast chromosomes. Aside from that, centromeres engage in relatively few long-range intrachromosomal interactions (Figure 3A). These findings have been confirmed in later studies of other budding yeasts, filamentous fungi and S. pombe [48–52]. Centromere clustering is such a prominent characteristic of yeast genome organization that the 3C signal has even been taken as a starting point to locate previously unidentified centromeres [53,54].

The remarkable plasticity of yeast genomes has allowed to drastically remodel the location and the number of centromeres. For example, selecting for neocentromeres in C. albicans, or manipulating the chromosome number by genome engineering in S. cerevisiae, result in a drastic reorganization of genomic contacts [55–57]. In both cases, upon the loss of a centromere, the strong interactions of flanking sequences with other centromeric regions disappear. In C. albicans, chromosomal regions near a neocentromere acquire new and strong interactions with all the other seven centromeric regions. The impact of this reorganization on gene expression was also analyzed because co-regulated genes often cluster along chromosome arms and in centromere-proximal regions [58,59]. In this case however, the impact on gene expression upon centromere delocalization is marginal, at least under standard laboratory growth conditions. In addition, reshuffling centromere location might also interfere with DNA metabolism and recombination. In fact, centromeres are among the earliest replicated regions in the yeast genome [48,60,61] and negatively impact the rate of cross-over during meiosis [62]. Future studies could therefore investigate the effects of centromere delocalization on other chromatin processes apart from gene expression.

Centromeric and pericentromeric regions also drive specific physical contacts in interphase of other organisms as visualized on their respective Hi-C contact maps (Figure 3). In line with their Rabl or non-Rabl organization, two different types of contact maps have been described in plants. The contact map of barley shows a strong cross shape signal within and between chromosomes, consistent with their typical Rabl conformation [63]. In contrast, the contact map of rice, tomato, foxtail millet and sorghum, that are described as non-Rabl, show a tripartite segregation of their chromosomes with centromeres and large pericentromeric heterochromatin grouped in a central B (inactive) compartment. Although interactions are enriched within all inactive compartments across the genome, they tend to be less involved
in interchromosomal interactions than the two distal A (active) compartments [32,64]. Though maize was previously reported to have an intermediate organization, Hi-C analysis demonstrated a strong intra- and interchromosomal cross signal, indicating a similar Rabl organization as in barley, at least in mesophyll protoplasts cells [32,64]. The genome-wide contact map of *A. thaliana* shows detectable interactions between all pairs of centromeric regions that make very little contacts with other genomic regions, except for the large heterochromatic locus on chromosome 4 called “knob” (Figure 3B) [65]. Here, centromere clustering gives rise to a less pronounced cross shape signal than observed in plants with Rabl conformation. Still, the strongest genome-wide Hi-C contacts of *A. thaliana* correspond to pericentromeric chromatin interactions, both within the same pericentromere and between different pericentromeres, consistent with the reported formation of heterochromatic chromocenters [65].

Centromere-centromere contacts are also visible on the *D. melanogaster* Hi-C matrix which shows strong interaction signals between the megabase long, heterochromatic, pericentromeric regions (Figure 3C) [66]. To explain these interactions, as well as other general chromosomal organization in flies (see below), it is hypothesized that (peri-)centromeric regions may assemble into their own nuclear compartment, again consistent with previous observations by microscopy [67].

Genome-wide Hi-C maps in vertebrates also show the presence of CTs, but centromere interactions are not as apparent (Figure 3D, E) [45,68]. The lack of strong centromere signals in vertebrate Hi-C contact maps is possibly due to weaker centromere interactions in these organisms. In addition, the lack of full centromere sequence assemblies might also be a confounding factor. Satellite sequences are systematically excluded from most next generation sequencing analysis due to their repetitive nature. Nevertheless, as the interaction signal spreads along neighboring chromatin fibers due to their physical properties, it is possible to recover the interaction encompassing repetitive regions by looking at adjacent non-ambiguous regions. Using this principle to reanalyze human and mouse Hi-C data, it was found that decomposing the contact map signal into a series of genomic tracks, known as eigenvectors, enable to determine the contribution of particular genomic features to higher-order chromosomal organization [69]. They further determine that the first eigenvector
relates to genomic sequences and local epigenetic chromatin states (as described before by
the compartmentalization), while the second (or third in the case of mouse) relates to the
position along the chromosome arm and highlights centromere-centromere interactions. The
association of all repetitive elements in the human and mouse genomes has also been studied
by using neighboring regions of ambiguous sequences, as well as stringent and restrictive
mapping parameters [70]. With these techniques, it was shown that some satellites
significantly co-localize inside the nucleus, particularly alpha-satellites in the human genome.
Furthermore, by using higher resolution maps of the human genome (down to 1 kb bins), the
two original compartments that were described by the first eigenvector were refined and
divided into five sub-compartments, two in A and three in B [71]. Sixty-two percent of
pericentromeric heterochromatin is found in sub-compartment B2, which is associated with
the nuclear lamina and nucleolus associated domains. Finally, superimposition of 3D genome
structure determination from single-cell Hi-C data with microscopy visualizing CENP-A reveal
a Rabl-conformation with centromeres and telomeres clustered at opposite sides of the
nucleus in G1 mouse ES cells [72]. Taken together, centromere clustering in vertebrates,
although less apparent in Hi-C contact maps, still account for a prevalent class of contacts that
have been demonstrated by many different means.

To overcome the problem of mapping repetitive sequences using a different strategy,
centromere interactions were analyzed in chicken cells, taking advantage of the non-repetitive
nature of centromeres on chromosomes 5, 27 and Z, as well as engineered strains with
neocentromeres at several positions on the Z chromosome [73]. As in humans,
eocentromeres in chicken are neither flanked by heterochromatin nor enriched for satellite
repeat sequences [74–76]. Using this set-up, it has been shown that repetitive and non-
repetitive centromeres associate with one another in 4C analyses and that non-repetitive
centromeres and neocentromeres interact with heterochromatic regions of the genome.

Centromere clustering even extends to Apicomplexans, single-cell organisms that
include the human malaria parasite *Plasmodium falciparum*. Recent Hi-C analyses of several
Plasmodium species and two additional apicomplexan parasites reveal varying degrees of
centromere and telomere clustering in most developmental stages. However, centromere
interactions appear to be lost in sporozoites, haploid cells that invade the insect salivary glands for transmission to the host [77,78].

To conclude, it is important to note that while the described features often arise from asynchronous cell populations they are unlikely the result of interactions coming from the low fraction of mitotic cells, as previously proposed in studies using synchronized cells [79–81]. Mitotic chromosomes are devoid of compartments and TADs and enriched in intrachromosomal long-range contacts compared to interphase chromosomes. This particular folding makes them drastically different from interphase chromosomes that resemble and better recapitulate chromosome conformation in exponentially growing cells. Taken together, despite different linear genome architectures and centromere sequence compositions, centromere interactions appear to be a common feature in many organisms scattered throughout the eukaryotic phylogeny.

Centromeres individualize chromosome arms

In addition to engaging in preferential interactions and building a specific sub-compartment, centromeres tend to create a barrier within each chromosome resulting in less frequent contacts between the chromosome arms on either side of them, compared to intra-arm contact frequencies (Figure 3F). This insulation property is visible on contact maps of species harboring a classical Rabl or Rabl-like chromosome conformation (Figure 3). Measurements in S. pombe for instance show more frequent intra-arm than inter-arm interactions; the latter being in the same range as interaction frequencies between chromosomes [50]. The barrier effect of centromeres also plays a role in species with other genome configurations. The first 3D studies in A. thaliana applying 4C on several viewpoints distributed across all five chromosomes demonstrate that chromosome arms are the main interaction units of the genome [82]. This result was later confirmed on genome-wide analysis showing a higher level of interactions between chromatin regions on the same side of a centromere, as opposed to interactions on opposite sides [65].

Genomic rearrangements involving the centromere in mutant flies demonstrate the insulation properties of centromeres. Interactions between polycomb target genes in D. melanogaster wildtype and in a mutant strain carrying a pericentromeric inversion (see
glossary) on chromosome 3 has been analyzed using 4C. This demonstrated preferential interactions between these loci when present on the same chromosome arm. These findings imply a strong rewiring of chromatin interactions in the rearranged strain consistent with a model of distinct territories formed by individual chromosome arms [67]. This model, however, has to be modulated depending on the developmental stage or tissue analyzed because relative interactions between arms and chromosomes are reported to vary [66,83,84].

Analyses in human cells using a modified 3C technique, called Tethered Conformation Capture (TCC), reveal that the contact profile of inactive regions decreases abruptly when located on opposite sides of the centromere [85]. This effect, however, is not seen for active regions which tend to be involved in long-range contacts regardless of the presence of the centromere [85]. These analyses demonstrate that the human centromere can also act as a contact barrier, at least for some categories of compartments.

In addition, the position of the centromere on the chromosome also constrains its overall architecture. This effect can be observed in S. cerevisiae where the small and long chromosome arms tend to interact more frequently with one another, respectively [86,87]. Furthermore, the position of the centromere on the chromosome also influences centromere-centromere interactions themselves. For instance, centromeric regions of human acrocentric chromosomes are more likely to contact each other than those of metacentric chromosomes [85].

Overall, the centromere influences interaction frequencies along the entire chromosome preventing certain contacts between particular chromosomal regions. Variation in the position of the centromere can directly impact long distance interactions and the general 3D conformation of chromosomes. It will be important to determine to what extent altered centromere locations can lead to perturbations in gene expression or other DNA related metabolisms in future studies.

**Concluding remarks and future perspectives**

In this review, we provide a collective view on the spatial organization of centromeres in a range of distinct eukaryotes (Key table). Despite their essential function, centromere
architectures and sequences are highly diverse. Yet, in many organisms the spatial organization of centromeres appears to be conserved. Centromeres localize to distinct nuclear sub-compartments or even cluster with one another in interphase nuclei. Given that the first layer of genome organization is the segregation of chromosomes into CTs with intrachromosomal interactions being the most frequent of all, centromere clustering across different chromosomes is not an intuitive result. This raises the questions of what leads to centromere clustering and whether there is a functional relevance to this organization (see outstanding question box).

Centromere clustering could be mediated by proteins that bind to centromeres or pericentromeres and stabilize long-range interactions between them. This mechanism could be similar to the formation of chromocenters in flies and mice that were proposed to be mediated by proteins bound to pericentromeric satellites and capable of bundling multiple DNA strands [88]. Future studies could aim to identify additional proteins involved in the formation of centromere clusters. The physical properties of centromere clusters could phase separate them into nuclear sub-compartments as it has been proposed to underlie A/B compartmentalization observed in Hi-C data [89] and the formation of heterochromatin foci bound by HP1 [90]. In fact, kinetochore components are enriched in coiled-coil and disordered domains [91] which represent structural features frequently associated with protein assemblies that induce protein phase separation [92].

Several observations point towards a functional relevance of centromere clustering. Recent studies have hypothesized that the aggregation of chromosomes via chromocenters in interphase could actually prevent micronuclei formation and thus play a role in genome stability [88]. This is consistent with observations showing that destabilization of spatial centromere positioning impairs silencing which in turn leads to an increase in transcription of transposable elements. This is associated with the accumulation of DNA double-strand breaks throughout the genome leading to mitotic defects and genome instability [35,93,94]. Additionally, centromere organization is subject to changes during the cell cycle and during various developmental stages including plant germination or floraison and mouse embryogenesis [95,96]. Several studies have also pointed out possible higher order cluster organization in human cells, with specific combinations of centromere associations
Hence, the question arises whether such non-random centromere associations have functional implications on surrounding chromatin processes. Still, given the dominant role of the cell cycle in determining genome architecture, temporal changes in cell-cycle stages will have to be taken into consideration when comparing centromere configurations across different organisms and developmental stages.

The position of the centromere on the chromosome together with its 3D organization also impacts the chromatin environment in its vicinity. It will therefore be interesting in future studies to investigate how genome architecture changes with centromere organization. This aspect could be addressed in organisms that have evolved neocentromeres or contain evolutionary new centromeres by comparing their genetic environment to the ancestral state. Additionally, studies in organisms with drastically different centromere organization, such as holocentric (see glossary) species, could also provide insights into the impact of centromeres on genome architecture.

Finally, a comprehensive picture on the conservation of 3D centromere conformation will require data from additional organisms including non-model organisms. However, in order to do this, we will have to overcome the challenges brought forth by the repetitive nature of many centromeres that result in incomplete genome assemblies. New and improved sequencing technologies will be required to get insights into centromere configurations in organisms with repetitive centromeres including humans.


**Box1: Centromere organization is diverse among eukaryotes.**

*S. cerevisiae* “point” centromeres are genetically defined by three genetic elements: CDEI, II and III and incorporate a single centromeric specific histone H3 variant, CenH3 (Cse4) [99,100].

In the Candida clade, including *C. albicans*, centromeres occur in large ORF-free regions of 4-18 kb which includes a CenH3-containing-nucleosome core region of 3-5 kb. They share no common sequence motif or repeats either within one organism or between them [5–7].

In *S. pombe*, the central core (cnt), of about 4 kb in length, assembles CenH3-containing-nucleosomes. It is surrounded by numerous alternating dg, dh and cen253 elements forming innermost (imr) and outermost (otr) repeats, giving rise to centromeric regions of 35–120 kb [101]. A similar organization with a non-repetitive mid core (2-5 kb) flanked by inverted repeats (2-5 kb) is found in the phylogenetically distant yeast *Candida tropicalis* [102].

In *Cryptococcus* species, centromeres are in syntenic positions in ORF-free, poorly transcribed regions featuring variable combinations of retrotransposons or their remnants. Centromeres span from 44 and 62 kb on average in *C. neoformans* and *C. deneoformans*, respectively, and about 14 kb in *C. deuterogatii* [103].

*A. thaliana* contains centromeres of several megabases with a central domain consisting mainly of 180 bp satellite DNA and fragments of Athila retrotransposon elements, 106B. The flanking regions contain multiple families of LTR retrotransposons and 5SrDNA sequences [15,16,104].

Rice and maize contain satDNA, named CentO (155 bp) and CentC (156 bp), respectively. Each also contains centromeric-specific retrotransposons, CRM (-Maize) and CRR (-Rice); while in wheat, CRW (-Wheat) are found, but no satDNA [14,17,18,105–108].

Analysis of the *D. melanogaster* centromere organization revealed that they are composed of islands of complex DNA enriched in retrotransposons flanked by arrays of short satellite repeats (5 to 12 bp) [19,20]. Whether satellite or retrotransposon sequences are the functional parts of *D. melanogaster* centromeres based on CENP-A enrichment remains however controversial [109].
Mouse chromosomes are all acrocentric, with centromeres composed of 120 bp minor-
satellites constituting the core centromeric region toward the telomere and the more
abundant 234 bp major-satellite forming pericentromeric heterochromatin [110,111].

The centromere of the human Y chromosome is made up of 171 bp alpha-satellite
monomers organized in a tandem fashion into large arrays of HOR [21]. While annotated
alpha-satellite arrays are dominated by HORs on BAC clones conferring chromosome-
specificity [112], another study proposes that the core of human centromeres is composed of
two classes of highly homogenous alpha-satellite dimers with HORs being pericentromeric
[113].
Figure 1: Sequence organization of centromeres in eukaryotes. For each species, the typical centromere organization is shown, with the size range for features of interest. When the centromere of a particular chromosome is depicted, its number is indicated after the species name and an approximate scale is shown on the right.
Figure 2. Schematic of centromere localization in interphase. Blue: DAPI, red: centromere probe by IF or FISH as indicated. Adapted from: *M. musculus*: [114]; *D. melanogaster*: [35]; *A. thaliana*: [34]; *T. aestivum* and *S. bicolor*: [32]; *C. neoformans*: [31]; *C. albicans*: [30]; *S. pombe*: [27]; *S. cerevisiae*: [115]. Scale bar = 5µm.
Figure 3. Normalized contact maps of (A) *S. cerevisiae*, full genome, resolution 5 kb, dataset from [86] (B) *A. thaliana*, full genome, resolution 20 kb, dataset from [65] (C) *D. melanogaster*, full genome, resolution 100 kb, dataset from [84]. (D) GM06990 human cell line, chr 3-5, resolution 1 Mb, dataset from [45]. (E) Chr 3 as in D; top half: normalized contact map; bottom half: correlation matrix (see glossary). Contact maps have been generated using the HiC-Pro suite, normalized by ICE using program default parameters [116] and drawn using HiCExplorer [117]. Correlation matrix has been generated by HiTC R package [118]. (F) Schematic of a simplified Hi-C contact map displaying three chromosomes. In the intrachromosomal area (black dashed box), chromosome arms are individualized on each side of the centromere. Centromere-centromere interactions are the main features observed in the interchromosomal area (white dashed box).
# Key table: Centromere organization in eukaryotes.

<table>
<thead>
<tr>
<th>Species</th>
<th>Nb of chr.</th>
<th>Type of centromere</th>
<th>Centromere imaging</th>
<th>Chromosome organization</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. cerevisiae</em></td>
<td>n = 16</td>
<td>Point centromeres</td>
<td>One focus near SPB</td>
<td>Rabl-like</td>
</tr>
<tr>
<td></td>
<td>2n = 32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. albicans</em></td>
<td>2n = 16</td>
<td>Regional without repetitive sequence</td>
<td>One focus near SPB</td>
<td>Rabl-like</td>
</tr>
<tr>
<td><em>S. pombe</em></td>
<td>n = 3</td>
<td>Regional with inverted repeats</td>
<td>One focus near SPB</td>
<td>Rabl-like</td>
</tr>
<tr>
<td></td>
<td>2n = 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. neoformans</em></td>
<td>n = 14</td>
<td>Regional with transposable elements</td>
<td>Individual foci at nuclear periphery</td>
<td>ND</td>
</tr>
<tr>
<td><em>A. thaliana</em></td>
<td>2n = 10</td>
<td>Regional with transposable elements</td>
<td>Clustered foci at chromocenters</td>
<td>Centromere clustering</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and SatDNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*O. sativa (rice) /</td>
<td>2n = 24 /</td>
<td>Regional with transposable elements</td>
<td>Dispersed foci in nucleus</td>
<td>Non-Rabl</td>
</tr>
<tr>
<td><em>S. bicolor</em></td>
<td>2n = 20</td>
<td>and SatDNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>T. aestivum</em> (wheat)</td>
<td>2n = 6X =42</td>
<td>ND</td>
<td>Clustered foci at one cell pole,</td>
<td>Rabl</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>opposite to nucleolus</td>
<td></td>
</tr>
<tr>
<td><em>D. melanogaster</em></td>
<td>2n = 8</td>
<td>Regional with transposable elements</td>
<td>Clustered foci around the nucleolus</td>
<td>Chromosome territory and arm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and SatDNA</td>
<td></td>
<td>territory</td>
</tr>
<tr>
<td><em>M. musculus</em></td>
<td>2n = 40</td>
<td>Regional with SatDNA</td>
<td>Clustered foci at chromocenters</td>
<td>Chromosome territory</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>localized at nuclear periphery</td>
<td></td>
</tr>
<tr>
<td><em>H. sapiens</em></td>
<td>2n = 46</td>
<td>Regional with SatDNA</td>
<td>Clustering at nuclear periphery</td>
<td>Chromosome territory</td>
</tr>
</tbody>
</table>

ND: not determined.
Glossary:

**Synteny:** conservation of the order of homologous genetic loci between two chromosomes.

**Neocentromeres:** new centromeres that form at an ectopic locus on the chromosome that is not centromeric. They typically arise due to the disruption of the endogenous centromere. Neocentromeres are often not associated with repetitive DNA but instead consist of unique sequences.

**Evolutionary new centromeres:** recently evolved novel centromeres appearing at ectopic locations on the chromosome compared to an ancestral centromere. Like neocentromeres, evolutionary new centromeres are often devoid of repetitive DNA.

**Chromocenter:** densely staining aggregation of heterochromatic regions in the nucleus of some cells.

**Chromosome Territory:** discrete region of the cell nucleus preferentially occupied by a particular chromosome.

**Chromosome Conformation Capture (3C):** method in molecular biology to analyze the spatial organization of chromatin in a cell between two specified loci.

**4C:** variation of 3C to visualize the contact pattern of one locus to any loci genome-wide.

**Hi-C:** variation of 3C in which biotin is incorporated at the ligation junction enabling it to be selected and obtain genome-wide contact maps.

**Contact map:** representation of a Hi-C experiment as a diagonally symmetrical square matrix indicating the frequency of contacts between any pairs of genomic loci.

**Pericentromeric Inversion:** an inversion is a chromosome rearrangement in which a segment of a chromosome is reversed end to end. In a pericentric inversions the segment includes the centromere.

**Holocentric:** referring to chromosomes that have multiple centromeres along their entire length, as opposed to monocentric where centromeres are restricted to one region on each chromosome.

**Correlation matrix:** normalized contact map where the Pearson correlation is taken between the $i^{th}$ row and $j^{th}$ column, which sharpens the plaid pattern.
References:


Jin, Q.W. et al. (2000) Centromere clustering is a major determinant of yeast interphase nuclear organization. *J. Cell Sci.* 113 (Pt 11), 1903–1912


of chromatin looping. Rao, S.S.P.

similar repetitive elements. Imakaev, M.

chromosomal translocations. Zhang, Y.

architecture. Tolhuis, B.

Drosophila g

Sexton, T.

cromatin interactions in Arabidopsis. Feng, S.

A/B Compartments. Dong, P.

gene. Mascher, M.

recombination during meiosis. Vincenten, N.

Raghuraman, M.K. Commun.

Centromere Cluster Burrack, L.S.

using Hi

Varoquaux, N. et al.


the diversity of chromosome organization in microorganisms. elife 3, e03318


Tolhuis, B. et al. (2011) Interactions among Polycomb domains are guided by chromosome architecture. PLoS Genet. 7, e1001343


24
Stevens, T.J. et al. (2017) 3D structures of individual mammalian genomes studied by single-cell
Hi-C. Nature 544, 59–64

Nishimura, K. et al. (2018) 3D genomic architecture reveals that neocentromeres associate with
heterochromatin regions. J. Cell Biol. DOI: 10.1083/jcb.201805003

Epigenetics Chromatin 3, 6

Hori, T. et al. (2014) Histone H4 Lys 20 monomethylation of the CENP-A nucleosome is essential

Shang, W.-H. et al. (2013) Chromosome engineering allows the efficient isolation of vertebrate
neocentromeres. Dev. Cell 24, 635–648

Ay, F. et al. (2014) Three-dimensional modeling of the P. falciparum genome during the
erthrocytic cycle reveals a strong connection between genome architecture and gene
expression. Genome Res. 24, 974–988


Lazar-Stefanita, L. et al. (2017) Cohesins and condensins orchestrate the 4D dynamics of yeast
chromosomes during the cell cycle. EMBO J. 36, 2684–2697


Grob, S. et al. (2013) Characterization of chromosomal architecture in Arabidopsis by
chromosome conformation capture. Genome Biol. 14, R129

Hou, C. et al. (2012) Gene density, transcription, and insulators contribute to the partition of the
Drosophila genome into physical domains. Mol. Cell 48, 471–484

Stadler, M.R. et al. (2017) Convergence of topological domain boundaries, insulators, and
polytene interbands revealed by high-resolution mapping of chromatin contacts in the early
Drosophila melanogaster embryo. eLife 6,

Kalhor, R. et al. (2011) Genome architectures revealed by tethered chromosome conformation

Mercy, G. et al. (2017) 3D organization of synthetic and scrambled chromosomes. Science 355,

Therizols, P. et al. (2010) Chromosome arm length and nuclear constraints determine the

Jagannathan, M. et al. (2018) A conserved function for pericentromeric satellite DNA. eLife 7,

Nuebler, J. et al. (2018) Chromatin organization by an interplay of loop extrusion and

Strom, A.R. et al. (2017) Phase separation drives heterochromatin domain formation. Nature 547,
241–245

of centromere-associated proteins from yeast to humans. Trends Cell Biol. 23, 260–269

Condensates. Trends Biochem. Sci. 43, 81–94


Mekhail, K. et al. (2008) Role for perinuclear chromosome tethering in maintenance of genome

Heterochromatin in Arabidopsis thaliana. Front. Plant Sci. 6, 1049

Martin, C. et al. (2006) Genome restructuring in mouse embryos during reprogramming and early

Ioannou, D. et al. (2017) A new model of sperm nuclear architecture following assessment of the


Choulet, F. et al. (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345, 1249721


Komissarov, A.S. et al. (2011) Tandemly repeated DNA families in the mouse genome. *BMC Genomics* 12, 531


