Trends in Genetics

CellPress
REVIEWS

Review

The Impact of Centromeres on Spatial Genome Architecture

Héloïse Muller,[1,2,*] José Gil, Jr,[1,2] and Ines Anna Drinnenberg[1,2]

The development of new technologies and experimental techniques is 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 chromatin regulation. Here we aim to share the structures and functions of centromeres in various 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.

Highlights

Centromeres' primary role is to ensure chromosome segregation. In addition to this well-known function, centromeres also have an impact on genome architecture.

Microscopy studies have shown a propensity of centromere to cluster in many organisms. In addition, these clusters are not randomly localized in the nuclear space and tend to be found close to the nuclear periphery or around the nucleolus.

At a large scale, chromosome architecture is constrained by centromeric interactions and the position of the centromere along each chromosome.

Chromosome conformation capture experiments confirmed centromere clustering while shedding light onto unexpected consequences for genome spatial regulation: centromeres act as strong topological barriers preventing specific types of contacts between the two chromosomal arms.

Centromeres: Beyond Chromosome Segregation

Eukaryotic genomes are not randomly organized. Chromatin is organized into multiple domains juxtaposed one after the other along the length of a chromosome and positioned into particular layouts and in 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. 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 nonrandom 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 the current limitations of and alternative strategies for 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. The length of centromeres ranges from as small as 120 bp to up to several megabases of DNA [4]. With some recent exceptions, sequence information is available only for shorter centromeres, whereas assemblies of larger centromeres are limited due to their length and repetitive sequence composition (Figure 1 and Box 1).

[1]Institut Curie, PSL Research University, CNRS, UMR3664, F-75005 Paris, France
[2]Sorbonne Université, Institut Curie, CNRS, UMR3664, F-75005 Paris, France

*Correspondence:
heloise.muller@curie.fr (H. Muller).

Centromere sequences of many yeasts and fungal species have been characterized (Figure 1 and Box 1). 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 its three centromeres, with the number of pericentromeric inverted repeats still unknown (Figure 1 and Box 1) (http://www.pombase.org). Compared with fungi, sequence assemblies of the megabase-long animal and plant centromeres, comprising 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 *Oryza sativa* and *Arabidopsis thaliana*, the molecular structure of large centromeric regions for several chromosomes has been determined (Figure 1 and 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 and 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 and 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 comprise chromosome-specific homogeneous arrays of tandem repeats, chromosomes 5, 27, and Z comprise 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** often form on sequences devoid of repetitive DNA [25] and tandem repeats have only recently been accumulating in some **evolutionarily new centromeres**. Regardless of their evolutionary framework, the nonrepetitive 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 *Saccharomyces 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 the nucleolus (Figure 2) [27–30]. Nevertheless, this organization can vary in some fungi. In *Cryptococcus neoformans*, 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 and non-Rabl. The Rabl organization is characterized by: (i) all centromeres being restricted to one pole of the nucleus, opposite the nucleolus; and (ii) 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]. By 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 reporting that chromosome organization can vary between different cell types. For example, most root

Glossary

4C: variation of 3C to visualize the contact pattern of one locus to any loci genome wide.
Chromocenter: densely staining aggregation of heterochromatic regions in the nucleus of some cells.
Chromosome conformation capture (3C): method in molecular biology to analyze the spatial organization of chromatin in a cell between two specified loci.
Chromosome territory (CT): discrete region of the cell nucleus preferentially occupied by a particular chromosome.
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.
Correlation matrix: normalized contact map where the Pearson correlation is taken between the *i*th row and the *j*th column, which sharpens the plaid pattern.
Evolutionarily new centromeres: recently evolved novel centromeres appearing at ectopic locations on the chromosome compared with an ancestral centromere. Like neocentromeres, evolutionarily new centromeres are often devoid of repetitive DNA.
Hi-C: variation of 3C in which biotin is incorporated at the ligation junction enabling it to be selected and to obtain genome-wide contact maps.
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.
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 comprise unique sequences.
Pericentromeric inversion: an inversion is a chromosome rearrangement in which a segment of a chromosome is reversed end to end. In a pericentric inversion, the segment includes the centromere.
Synteny: conservation of the order of homologous genetic loci between two chromosomes.

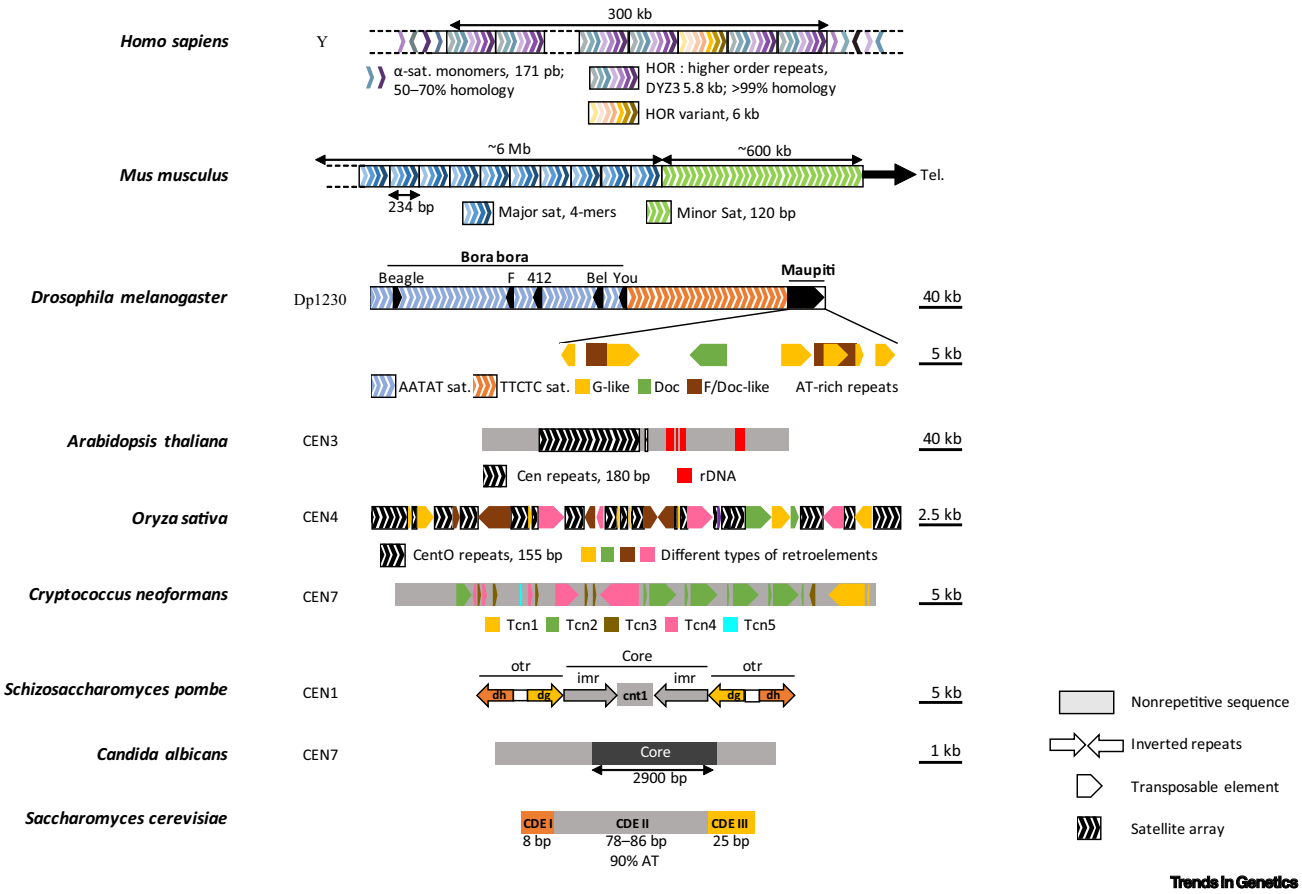

Figure 1. Sequence Organization of Centromeres in Eukaryotes. For each species, the typical centromere organization is shown, with the size range for features of interest. When the centromere of a particular chromosome is depicted, its number is indicated after the species name and an approximate scale is shown on the right.

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 with the Rabl and non-Rabl conformations, *A. thaliana* has a distinct centromeric organization with (peri-)centromeric heterochromatin organized as clearly distinguishable **chromocenters** 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) fewer than ten (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 four to six clusters (Figure 2) and third-instar larval diploid ($2n = 8$) hemocytes show two or three 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

Box 1. Centromere Organization Is Diverse among Eukaryotes

Saccharomyces cerevisiae 'point' centromeres are genetically defined by three genetic elements – CDEI, II, and III – and incorporate a single centromere-specific histone H3 variant, CenH3 (Cse4) [99,100].

In the *Candida* clade, including *Candida albicans*, centromeres occur in large open reading frame (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 *Schizosaccharomyces 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 nonrepetitive 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 to 62 kb on average in *C. neoformans* and *C. deneoformans*, respectively, and about 14 kb in *Candida deuterogatii* [103].

Arabidopsis thaliana contains centromeres of several megabases with a central domain comprising mainly 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 the centromere-specific retrotransposons CRM (maize) and CRR (rice), while in wheat CRW (wheat) is found but no SatDNA [14,17,18,105–108].

Analysis of the *Drosophila melanogaster* centromere organization revealed that they comprise islands of complex DNA enriched in retrotransposons flanked by arrays of short satellite repeats (5–12 bp) [19,20]. Whether satellite or retrotransposon sequences are the functional parts of *D. melanogaster* centromeres based on CenH3 (Cid) enrichment remains, however, controversial [109].

Mouse chromosomes are all acrocentric, with centromeres comprising 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 comprises 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 comprises two classes of highly homogeneous alpha-satellite dimers with HORs being pericentromeric [113].

chromocenters comprising pericentromeric heterochromatin (Figure 2) [36,37]. More recent analyses of nondividing/quiescent (G0) lymphocytes from mouse and humans revealed that centromeres cluster preferentially at the nuclear periphery, forming on average 13 or nine signals per human ($2n = 46$) or mouse ($2n = 40$) nucleus, respectively [38]. In the few cases of 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)**], centromeres are found at the periphery of CTs in 95% of studied cases and positioned toward 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. On exit from mitosis, in early G1, very little to no clustering is observed, and the number of kinetochore signals decrease in late G1 and S phase 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 is less pronounced in the nuclei

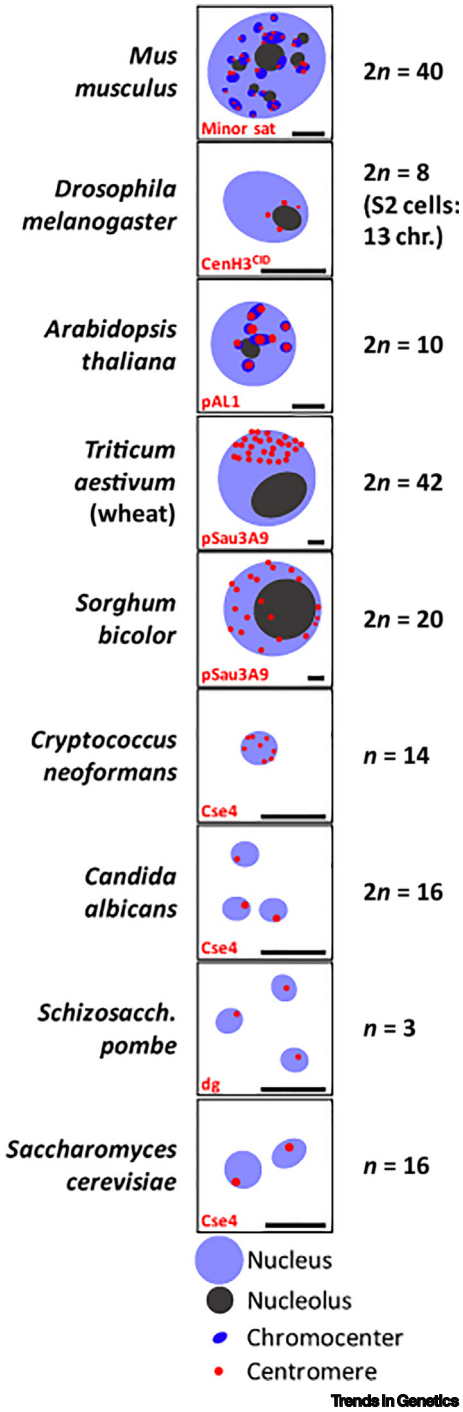

Figure 2. Schematic of Centromere Localization in Interphase. Blue, DAPI; red, centromere probe by immunofluorescence (IF) or fluorescence *in situ* hybridization (FISH) as indicated. Adapted from: *Mus musculus* [114]; *Drosophila melanogaster* [35]; *Arabidopsis thaliana* [34]; *Triticum aestivum* and *Sorghum bicolor* [32]; *Cryptococcus neoformans* [31]; *Candida albicans* [30]; *Schizosaccharomyces pombe* [27]; *Saccharomyces cerevisiae* [115]. Bar, 5 μm.

of cycling cells or progenitor cells compared with noncycling 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)** and its derivatives (**4C**, **Hi-C**, etc.) offer new ways to study the importance of centromeres in genome organization. Early studies define the top of the genome organization hierarchy as the CTs [44]. In line with that, Hi-C data from various organisms demonstrate 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, and 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 that 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 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 of previously described centromere interactions visible on Hi-C matrices: (i) centromeres build a specific subcompartment inside the nucleus; and (ii) they form a barrier to intrachromosomal arm interactions.

Centromeres Form Strong Interacting Subcompartments

3D genome organization has been studied in several fungal species. The first *S. cerevisiae* full-genome **contact map** 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 Rabl-like organization of budding yeast chromosomes. Besides 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 drastic remodeling of the location and number of centromeres. For example, selecting for neocentromeres in *Candida albicans* or manipulating the chromosome number by genome engineering in *S. cerevisiae* results 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 of the other seven centromeric regions. The impact of this reorganization on gene expression was also analyzed because coregulated 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, the reshuffling of centromere location might also interfere with DNA metabolism and recombination. Centromeres are among the earliest-replicated regions in the yeast genome [48,60,61] and negatively impact the rate of crossover during meiosis [62]. Future

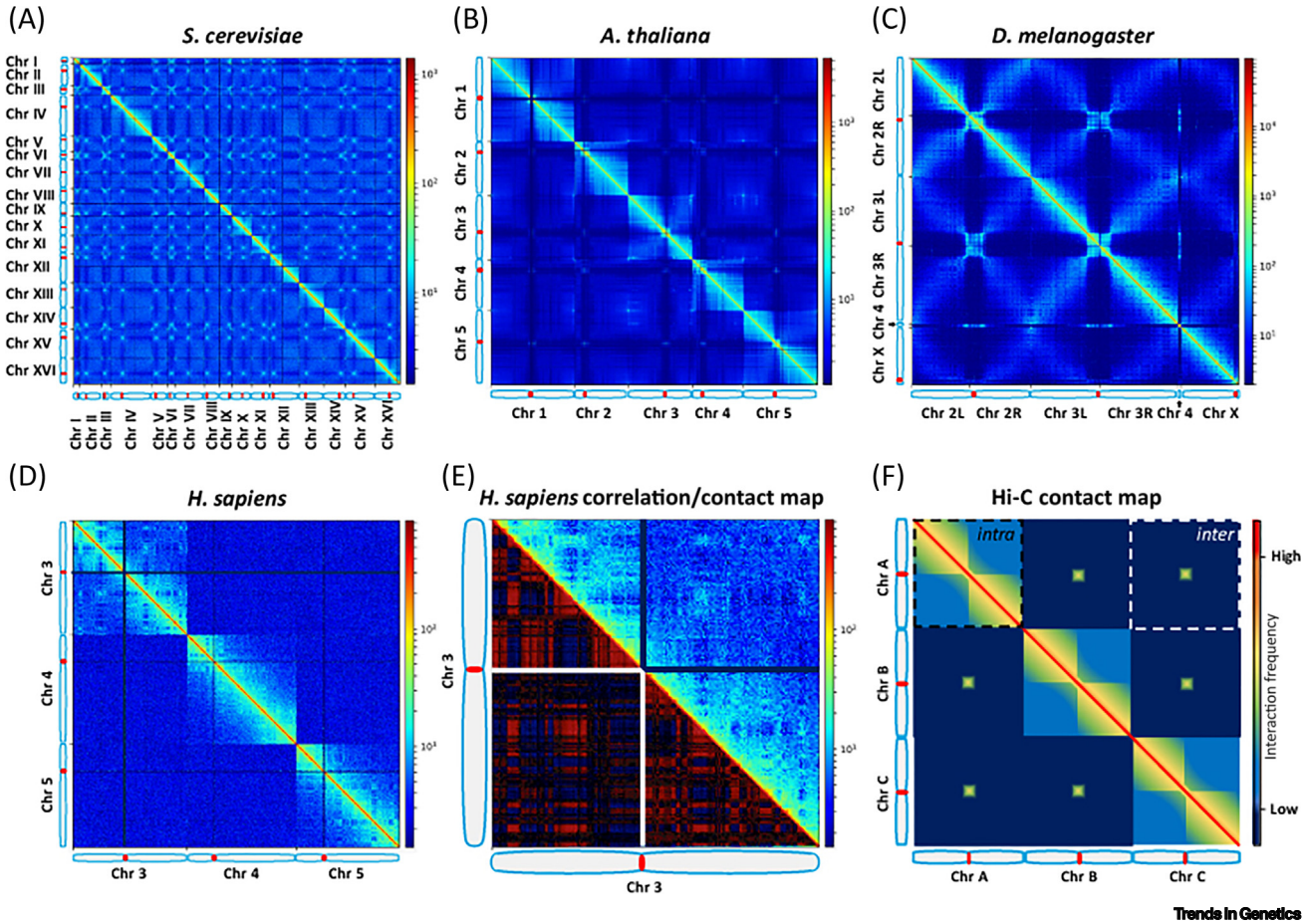

Figure 3. Normalized Contact Maps. (A) *Saccharomyces cerevisiae*: full genome, resolution 5 kb, dataset from [86]. (B) *Arabidopsis thaliana*: full genome, resolution 20 kb, dataset from [65]. (C) *Drosophila melanogaster*: full genome, resolution 100 kb, dataset from [84]. (D) GM06990 human cell line: chromosome 3–5, resolution 1 Mb, dataset from [45]. (E) Chromosome 3 as in (D): top half, normalized contact map; bottom half, **correlation matrix**. Contact maps have been generated using the HiC-Pro suite, normalized by ICE using program default parameters [116] and drawn using HiCExplorer [117]. The correlation matrix has been generated by the HiTC R package [118]. (F) Schematic of a simplified Hi-C contact map displaying three chromosomes. In the intrachromosomal area (black-broken-line box), chromosome arms are individualized on each side of the centromere. Centromere–centromere interactions are the main features observed in the interchromosomal area (white-broken-line box).

studies could therefore investigate the effects of centromere delocalization on other chromatin processes besides gene expression.

Centromeric and pericentromeric regions also drive specific physical contacts in interphase in 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 map have been described in plants. The contact map of barley shows a strong cross-shaped signal within and between chromosomes, consistent with their typical Rabl conformation [63]. By contrast, the contact map of rice, tomato, foxtail millet, and sorghum, which are described as non-Rabl, show 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]. Although maize was previously reported to have an intermediate organization, Hi-C analysis demonstrated a strong intra- and interchromosomal cross signal, indicating a Rabl organization similar to that 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 few 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-shaped 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 *Drosophila 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 be a confounding factor. Satellite sequences are systematically excluded from most next-generation sequencing analyses 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 nonambiguous 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, enabled the determination of the contribution of particular genomic features to higher-order chromosomal organization [69]. They further determined 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 the 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 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 colocalize 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 subcompartments, two in A and three in B [71]. Sixty-two percent of pericentromeric heterochromatin is found in subcompartment 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]. Taking these findings 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 nonrepetitive 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, neocentromeres in the chicken are neither flanked by heterochromatin nor enriched for satellite repeat sequences [74–76]. Using this setup, it has been shown that repetitive and nonrepetitive centromeres associate with one

another in 4C analyses and that nonrepetitive centromeres and neocentromeres interact with heterochromatic regions of the genome.

Centromere clustering even extends to apicomplexans, single-celled 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 to be 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 with interphase chromosomes. This particular folding makes them drastically different from interphase chromosomes that resemble and better recapitulate the chromosome conformation in exponentially growing cells. Taking these findings 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 subcompartment, centromeres tend to create a barrier within each chromosome resulting in less frequent contacts between the chromosome arms on either side of them compared with 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 interarm 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 than interactions on opposite sides [65].

Genomic rearrangements involving the centromere in mutant flies demonstrate the insulating properties of centromeres. Interactions between polycomb target genes in *D. melanogaster* wild type and in a mutant strain carrying a **pericentromeric inversion** 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 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) revealed that the contact profile of inactive regions decreases abruptly when they are 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 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 influences centromere–centromere interactions themselves. For instance, the 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 extend altered centromere locations can lead to perturbations in gene expression or other DNA-related metabolism 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 (Table 1, 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 subcompartments 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 Questions).

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 subcompartments as has been proposed to underlie A/B compartmentalization observed in Hi-C data [89] and the formation of heterochromatin foci bound by HP1 [90]. 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 toward a functional relevance of centromere clustering. Recent studies have hypothesized that the aggregation of chromosomes via chromocenters in interphase could prevent micronucleus formation and thus play a role in genome stability [88]. This is consistent with observations showing that the destabilization of spatial centromere positioning impairs silencing, which in turn leads to an increase in the 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 [39,40,85,97,98]. Hence, the question arises of whether such nonrandom centromere associations have functional implications for surrounding chromatin processes. However, 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.

Outstanding Questions

How are centromeres organized and what is the identity of centromeric sequences?

What leads to centromere clustering and which factors are involved?

In particular, how is centromere clustering affected on the depletion of factors affecting chromatin organization and chromocenter formation, including DNA methyltransferases and methyl CpG-binding domain proteins, histone-modifying enzymes, and known architectural proteins?

How are centromeres localized within the nuclear space?

In particular, which factors enable centromere localization to specific nuclear compartments, including the nuclear lamina and the nucleolus?

What are the consequences of centromere interactions for genome architecture?

To what extent can the information on spatial genome architecture explain and predict the location of neocentromere formation?

Is there a functional relevance of centromere clustering for DNA metabolism, including gene expression or replication?

How do changes in centromere organization or compaction (i.e. developmental, malignant, or evolutionary changes) lead to changes in 3D nuclear organization and what is the effect?

Key Table

Table 1. Centromere Organization in Eukaryotes

Species	Number of chromosomes	Type of centromere	Centromere imaging	Chromosome organization
Saccharomyces cerevisiae	*n* = 16 2*n* = 32	Point centromeres	One focus near SPB	Rabl-like
Candida albicans	2*n* = 16	Regional without repetitive sequence	One focus near SPB	Rabl-like
Schizosaccharomyces pombe	*n* = 3 2*n* = 6	Regional with inverted repeats	One focus near SPB	Rabl-like
Cryptococcus neoformans	*n* = 14	Regional with transposable elements	Individual foci at nuclear periphery	ND
Arabidopsis thaliana	2*n* = 10	Regional with transposable elements and SatDNA	Clustered foci at chromocenters	Centromere clustering
Oryza sativa (rice)/ *Sorghum bicolor*	2*n* = 24/2*n* = 20	Regional with transposable elements and SatDNA	Dispersed foci in nucleus	Non-Rabl
Triticum aestivum (wheat)	2*n* = 6X = 42	ND	Clustered foci at one cell pole, opposite nucleolus	Rabl
Drosophila melanogaster	2*n* = 8	Regional with transposable elements and SatDNA	Clustered foci around the nucleolus	CT and arm territory
Mus musculus	2*n* = 40	Regional with SatDNA	Clustered foci at chromocenters localized at nuclear periphery	CT
Homo sapiens	2*n* = 46	Regional with SatDNA	Clustering at nuclear periphery	CT

ND, not determined.

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 evolutionarily new centromeres by comparing their genetic environment with the ancestral state. Additionally, studies in organisms with drastically differing centromere organization, such as **holocentric** 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, 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 obtain insights into centromere configurations in organisms with repetitive centromeres, including humans.

Acknowledgment

We thank Harmit Malik and all the members of the Drinnenberg lab for discussion. I.A.D. 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, the Institut Curie and funds from the Atip Avenir 2015 program and the ERC (758757-CENEVO).

References

1. Rowley, M.J. and Corces, V.G. (2018) Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800
2. Towbin, B.D. *et al.* (2013) Mechanisms of heterochromatin subnuclear localization. *Trends Biochem. Sci.* 38, 356–363
3. Flemming, W. (1882) *Zellsubstanz, Kern und Zelltheilung, F.C.W. Vogel (in German)*
4. Steiner, F.A. and Henikoff, S. (2015) Diversity in the organization of centromeric chromatin. *Curr. Opin. Genet. Dev.* 31, 28–35
5. Sanyal, K. *et al.* (2004) Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11374–11379

6. Mishra, P.K. *et al.* (2007) Centromere size and position in *Candida albicans* are evolutionarily conserved independent of DNA sequence heterogeneity. *Mol. Gen. Genomics.* 278, 455–465
7. Padmanabhan, S. *et al.* (2008) Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, *Candida albicans* and *Candida dubliniensis*. *Proc. Natl. Acad. Sci. U. S. A.* 105, 19797–19802
8. Dujon, B. *et al.* (2004) Genome evolution in yeasts. *Nature* 430, 35–44
9. Gordon, J.L. *et al.* (2011) Mechanisms of chromosome number evolution in yeast. *PLoS Genet.* 7, e1002190
10. Meraldi, P. *et al.* (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol.* 7, R23
11. Chin, C-S. *et al.* (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054
12. Michael, T.P. *et al.* (2018) High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* 9, 541
13. Solares, E.A. *et al.* (2018) Rapid low-cost assembly of the *Drosophila melanogaster* reference genome using low-coverage, long-read sequencing. *G3 (Bethesda)* 8, 3143–3154
14. Cheng, Z. *et al.* (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* 14, 1691–1704
15. Kumekawa, N. *et al.* (2000) The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 5. *DNA Res.* 7, 315–321
16. Kumekawa, N. *et al.* (2001) The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 4. *DNA Res.* 8, 285–290
17. Wu, J. *et al.* (2004) Composition and structure of the centromeric region of rice chromosome 8. *Plant Cell* 16, 967–976
18. Zhang, Y. *et al.* (2004) Structural features of the rice chromosome 4 centromere. *Nucleic Acids Res.* 32, 2023–2030
19. Chang, C.H. *et al.* (2019) Islands of retroelements are major components of *Drosophila centromeres*. *PLoS Biol.* 17, e3000241
20. Sun, X. *et al.* (2003) Sequence analysis of a functional *Drosophila* centromere. *Genome Res.* 13, 182–194
21. Jain, M. *et al.* (2018) Linear assembly of a human centromere on the Y chromosome. *Nat. Biotechnol.* 36, 321–323
22. Gong, Z. *et al.* (2012) Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *Plant Cell* 24, 3559–3574
23. Piras, F.M. *et al.* (2010) Uncoupling of satellite DNA and centromeric function in the genus *Equus*. *PLoS Genet.* 6, e1000845
24. Shang, W-H. *et al.* (2010) Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res.* 20, 1219–1228
25. Nergadze, S.G. *et al.* (2018) Birth, evolution, and transmission of satellite-free mammalian centromeric domains. *Genome Res.* 28, 789–799
26. Musacchio, A. and Desai, A. (2017) A molecular view of kinetochore assembly and function. *Biology (Basel)* 6, E5
27. Funabiki, H. *et al.* (1993) Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *J. Cell Biol.* 121, 961–976
28. Guacci, V. *et al.* (1997) Centromere position in budding yeast: evidence for anaphase A. *Mol. Biol. Cell* 8, 957–972
29. Jin, Q.W. *et al.* (2000) Centromere clustering is a major determinant of yeast interphase nuclear organization. *J. Cell Sci.* 113, 1903–1912
30. Thakur, J. and Sanyal, K. (2012) A coordinated interdependent protein circuitry stabilizes the kinetochore ensemble to protect CENP-A in the human pathogenic yeast *Candida albicans*. *PLoS Genet.* 8, e1002661
31. Kozubowski, L. *et al.* (2013) Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast *Cryptococcus neoformans*. *MBio* 4, e00614-13
32. Dong, F. and Jiang, J. (1998) Non-Rabl patterns of centromere and telomere distribution in the interphase nuclei of plant cells. *Chromosom. Res.* 6, 551–558
33. Prieto, P. *et al.* (2004) Chromosomes associate premeiotically and in xylem vessel cells via their telomeres and centromeres in diploid rice (*Oryza sativa*). *Chromosoma* 112, 300–307
34. Fransz, P. *et al.* (2002) Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14584–14589
35. Padeken, J. *et al.* (2013) The nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus. *Mol. Cell* 50, 236–249
36. Brown, K.E. *et al.* (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* 91, 845–854
37. Vourc'h, C. *et al.* (1993) Cell cycle-dependent distribution of telomeres, centromeres, and chromosome-specific subsatellite domains in the interphase nucleus of mouse lymphocytes. *Exp. Cell Res.* 205, 142–151
38. Weierich, C. *et al.* (2003) Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. *Chromosom. Res.* 11, 485–502
39. Alcobia, I. *et al.* (2003) The spatial organization of centromeric heterochromatin during normal human lymphopoiesis: evidence for ontogenically determined spatial patterns. *Exp. Cell Res.* 290, 358–369
40. Alcobia, I. *et al.* (2000) Spatial associations of centromeres in the nuclei of hematopoietic cells: evidence for cell-type-specific organizational patterns. *Blood* 95, 1608–1615
41. Bartholdi, M.F. (1991) Nuclear distribution of centromeres during the cell cycle of human diploid fibroblasts. *J. Cell Sci.* 99, 255–263
42. Solovei, I. *et al.* (2004) Differences in centromere positioning of cycling and postmitotic human cell types. *Chromosoma* 112, 410–423
43. Wiblin, A.E. (2005) Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. *J. Cell Sci.* 118, 3861–3868
44. Cremer, T. and Cremer, M. (2010) Chromosome territories. *Cold Spring Harb. Perspect. Biol.* 2, a003889
45. Lieberman-Aiden, E. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293
46. Dekker, J. and Mirny, L. (2016) The 3D genome as moderator of chromosomal communication. *Cell* 164, 1110–1121
47. Duan, Z. *et al.* (2010) A three-dimensional model of the yeast genome. *Nature* 465, 363–367
48. Descorps-Declère, S. *et al.* (2015) Genome-wide replication landscape of *Candida glabrata*. *BMC Biol.* 13, 69
49. Galazka, J.M. *et al.* (2016) Neurospora chromosomes are organized by blocks of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. *Genome Res.* 26, 1069–1080
50. Kakui, Y. *et al.* (2017) Condensin-mediated remodeling of the mitotic chromatin landscape in fission yeast. *Nat. Genet.* 49, 1553–1557
51. Marbouty, M. *et al.* (2014) Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *Elife* 3, e03318
52. Mizuguchi, T. *et al.* (2014) Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* 516, 432–435
53. Marie-Nelly, H. *et al.* (2014) Filling annotation gaps in yeast genomes using genome-wide contact maps. *Bioinformatics* 30, 2105–2113
54. Varoquaux, N. *et al.* (2015) Accurate identification of centromere locations in yeast genomes using Hi-C. *Nucleic Acids Res.* 43, 5331–5339
55. Burrack, L.S. *et al.* (2016) Neocentromeres provide chromosome segregation accuracy and centromere clustering to multiple loci along a *Candida albicans* chromosome. *PLoS Genet.* 12, e1006317
56. Luo, J. *et al.* (2018) Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* 560, 392–396
57. Shao, Y. *et al.* (2018) Creating a functional single-chromosome yeast. *Nature* 560, 331–335

58. Cohen, B.A. *et al.* (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* 26, 183–186
59. Janga, S.C. *et al.* (2008) Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 105, 15761–15766
60. Agier, N. *et al.* (2018) The evolution of the temporal program of genome replication. *Nat. Commun.* 9, 2199
61. Raghuraman, M.K. *et al.* (2001) Replication dynamics of the yeast genome. *Science* 294, 115–121
62. Vincenten, N. *et al.* (2015) The kinetochore prevents centromere-proximal crossover recombination during meiosis. *Elife* 4, e10850
63. Mascher, M. *et al.* (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433
64. Dong, P. *et al.* (2017) 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant* 10, 1497–1509
65. Feng, S. *et al.* (2014) Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in *Arabidopsis*. *Mol. Cell* 55, 694–707
66. Sexton, T. *et al.* (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472
67. Tolhuis, B. *et al.* (2011) Interactions among polycomb domains are guided by chromosome architecture. *PLoS Genet.* 7, e1001343
68. Zhang, Y. *et al.* (2012) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148, 908–921
69. Imakaev, M. *et al.* (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* 9, 999–1003
70. Cournac, A. *et al.* (2016) The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res.* 44, 245–255
71. Rao, S.S.P. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680
72. Stevens, T.J. *et al.* (2017) 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544, 59–64
73. Nishimura, K. *et al.* (2018) 3D genomic architecture reveals that neocentromeres associate with heterochromatin regions. *J. Cell Biol.* 218, 134–149
74. Alonso, A. *et al.* (2010) A paucity of heterochromatin at functional human neocentromeres. *Epigenetics Chromatin* 3, 6
75. Hori, T. *et al.* (2014) Histone H4 Lys 20 monomethylation of the CENP-A nucleosome is essential for kinetochore assembly. *Dev. Cell* 29, 740–749
76. Shang, W-H. *et al.* (2013) Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. *Dev. Cell* 24, 635–648
77. Ay, F. *et al.* (2014) Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.* 24, 974–988
78. Bunnik, E.M. *et al.* (2019) Comparative 3D genome organization in apicomplexan parasites. *Proc. Natl. Acad. Sci. U. S. A.* 116, 3183–3192
79. Gibcus, J.H. *et al.* (2018) A pathway for mitotic chromosome formation. *Science* 359, eaao6135
80. 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
81. Naumova, N. *et al.* (2013) Organization of the mitotic chromosome. *Science* 342, 948–953
82. Grob, S. *et al.* (2013) Characterization of chromosomal architecture in *Arabidopsis* by chromosome conformation capture. *Genome Biol.* 14, R129
83. 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
84. 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, e29550
85. Kalhor, R. *et al.* (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* 30, 90–98
86. Mercy, G. *et al.* (2017) 3D organization of synthetic and scrambled chromosomes. *Science* 355, eaaf4597
87. Therizols, P. *et al.* (2010) Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci. U. S. A.* 107, 2025–2030
88. Jagannathan, M. *et al.* (2018) A conserved function for pericentromeric satellite DNA. *Elife* 7, e34122
89. Nuebler, J. *et al.* (2018) Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U. S. A.* 115, E6697–E6706
90. Strom, A.R. *et al.* (2017) Phase separation drives heterochromatin domain formation. *Nature* 547, 241–245
91. Westermann, S. and Schleiffer, A. (2013) Family matters: structural and functional conservation of centromere-associated proteins from yeast to humans. *Trends Cell Biol.* 23, 260–269
92. Woodruff, J.B. *et al.* (2018) Organization and function of non-dynamic biomolecular condensates. *Trends Biochem. Sci.* 43, 81–94
93. Peng, J.C. and Karpen, G.H. (2007) H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat. Cell Biol.* 9, 25–35
94. Mekhail, K. *et al.* (2008) Role for perinuclear chromosome tethering in maintenance of genome stability. *Nature* 456, 667–670
95. Simon, L. *et al.* (2015) Structure and function of centromeric and pericentromeric heterochromatin in *Arabidopsis thaliana*. *Front. Plant Sci.* 6, 1049
96. Martin, C. *et al.* (2006) Genome restructuring in mouse embryos during reprogramming and early development. *Dev. Biol.* 292, 317–332
97. Ioannou, D. *et al.* (2017) A new model of sperm nuclear architecture following assessment of the organization of centromeres and telomeres in three-dimensions. *Sci. Rep.* 7, 41585
98. Tjong, H. *et al.* (2016) Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc. Natl. Acad. Sci. U. S. A.* 113, E1663–E1672
99. Fitzgerald-Hayes, M. *et al.* (1982) Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell* 29, 235–244
100. Furuyama, S. and Biggins, S. (2007) Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* 104, 14706–14711
101. Wood, V. *et al.* (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415, 871–880
102. Chatterjee, G. *et al.* (2016) Repeat-associated fission yeast-like regional centromeres in the ascomycetous budding yeast *Candida tropicalis*. *PLoS Genet.* 12, e1005839
103. Yadav, V. *et al.* (2018) RNAi is a critical determinant of centromere evolution in closely related fungi. *Proc. Natl. Acad. Sci. U. S. A.* 115, 3108–3113
104. Initiative, Arabidopsis Genome (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
105. Choulet, F. *et al.* (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345, 1249721
106. International Wheat Genome Sequencing Consortium (IWGSC) (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345, 1251788
107. Schnable, P.S. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115
108. Wolfgruber, T.K. *et al.* (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. *PLoS Genet.* 5, e1000743
109. Talbert, P.B. *et al.* (2018) Simple and complex centromeric satellites in *Drosophila* sibling species. *Genetics* 208, 977–990
110. Komissarov, A.S. *et al.* (2011) Tandemly repeated DNA families in the mouse genome. *BMC Genomics* 12, 531
111. Kuznetsova, I. *et al.* (2006) High-resolution organization of mouse centromeric and pericentromeric DNA. *Cytogenet. Genome Res.* 112, 248–255

112. Aldrup-Macdonald, M.E. and Sullivan, B.A. (2014) The past, present, and future of human centromere genomics. *Genes* 5, 33–50
113. Henikoff, J.G. *et al.* (2015) A unique chromatin complex occupies young α-satellite arrays of human centromeres. *Sci. Adv.* 1, e1400234
114. Markaki, Y. *et al.* (2013) Fluorescence *in situ* hybridization applications for super-resolution 3D structured illumination microscopy. *Methods Mol. Biol.* 950, 43–64
115. Chen, Y. *et al.* (2000) The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. *Mol. Cell. Biol.* 20, 7037–7048
116. Servant, N. *et al.* (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259
117. Wolff, J. *et al.* (2018) Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* 46, W11–W16
118. Servant, N. *et al.* (2012) HiTC: exploration of high-throughput "C" experiments. *Bioinformatics* 28, 2843–2844