genes

MDPI

Review

The Dynamic Structure and Rapid Evolution of Human Centromeric Satellite DNA

Glennis A. Logsdon [1] and Evan E. Eichler [1,2,*]

[1] Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA
[2] Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA
[*] Correspondence: eee@gs.washington.edu; Tel.: +1-(206)-543-9526

Abstract: The complete sequence of a human genome provided our first comprehensive view of the organization of satellite DNA associated with heterochromatin. We review how our understanding of the genetic architecture and epigenetic properties of human centromeric DNA have advanced as a result. Preliminary studies of human and nonhuman ape centromeres reveal complex, saltatory mutational changes organized around distinct evolutionary layers. Pockets of regional hypomethylation within higher-order α-satellite DNA, termed centromere dip regions, appear to define the site of kinetochore attachment in all human chromosomes, although such epigenetic features can vary even within the same chromosome. Sequence resolution of satellite DNA is providing new insights into centromeric function with potential implications for improving our understanding of human biology and health.

Keywords: centromere; pericentromeric DNA; satellite DNA; genomics; long-read sequencing

check for updates

Citation: Logsdon, G.A.; Eichler, E.E. The Dynamic Structure and Rapid Evolution of Human Centromeric Satellite DNA. *Genes* **2023**, *14*, 92. https://doi.org/10.3390/genes14010092

Academic Editors: Manuel A. Garrido-Ramos, Miroslav Plohl and Eva Šatović-Vukšić

Received: 15 December 2022
Revised: 22 December 2022
Accepted: 24 December 2022
Published: 28 December 2022

1. Introduction

Centromeres are essential chromosomal regions that serve as the site for spindle attachment during mitosis and meiosis and ensure the equal and accurate segregation of chromosomes during cell division. In almost all mammalian species, centromeres are composed of arrays of near-identical tandem repeats, which were identified in early human DNA centrifugation experiments as AT-rich DNA fractions termed "satellite DNA" [1–3]. For the past two decades, the majority of centromeric satellite DNA have been excluded from human reference genomes and have been, instead, represented as unfinished sequence gaps or simulated arrays known as "reference models" or "decoys" [4]. In addition, most sequencing-based studies have excluded these regions as part of standard human genetic analyses. Consequently, our understanding of the sequence organization, variation, and evolution of human centromeres has been limited, owing to their highly repetitive nature, large size (often megabases in length), and high sequence identity.

A series of pioneering studies based on pulsed-field gel electrophoresis, Southern blotting, and fluorescence *in situ* hybridizations in the late 1980s and 1990s revealed much about the organization of human centromeric satellite DNA [5–11]. Human centromeres are mainly composed of six classes of repetitive DNA: α-satellite, β-satellite, γ-satellite, and three shorter motifs termed HSATI, HSATII, and HSATIII (Figure 1, Table 1). While α-satellites are found on every chromosome in association with the primary metaphase constriction [12–15], the other satellites are restricted to a subset of chromosomes [16–20], chromosomal regions [21–25], or secondary constrictions [26–28], also called qh regions (Table 1). The kinetochore is largely restricted to higher-order repeat (HOR) units of the tandemly repeating α-satellite, which are flanked on the periphery by more divergent monomeric α-satellite DNA followed by other classes of satellites. While the chromosome-specific α-satellite HORs in humans were all defined by the mid-1990s, only general models existed for how centromeres are organized and have evolved [29], with a limited understanding of their precise sequence composition and how they vary among human and nonhuman primates [30–38]. Moreover, the site of kinetochore attachment was not clearly defined, often being inferred based on lower-resolution cytogenetic and immunohistochemical experiments or through limited sequence analyses [39–41]. With the application of long-read sequencing technologies, centromeric satellites can now be fully sequenced and assembled [42–45]. We review the complete sequence of human centromeric satellite DNA with an emphasis on the new insights that have emerged and how our model of centromeric DNA has been further refined, with potential implications for improving our understanding of human biology and health.

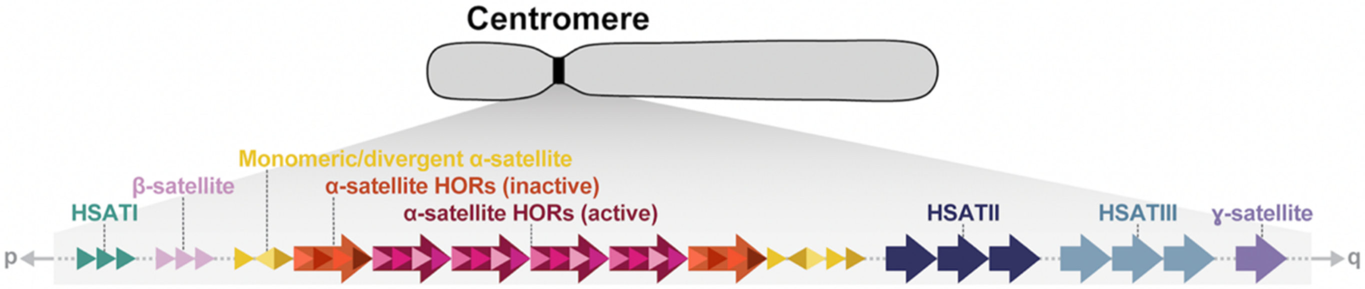
Centromere
Monomeric/divergent α-satellite
β-satellite
α-satellite HORs (inactive)
HSATI
α-satellite HORs (active)
HSATII
HSATIII
γ-satellite
p
q

Figure 1. Model of human centromeric and pericentromeric regions. Schematic of the generalized organization of human centromeres and their flanking sequence. Major components and their structures are shown. HORs, higher-order repeats; HSAT, human satellite. Not drawn to scale.

Table 1. Composition and abundance of repeats within human centromeric and pericentromeric regions.

Peri/Centromere Repeat Type	**Repeat Unit Length (bp)**	**GC Content (%)**	**Abundance inthe T2T-CHM13 Genome (Mbp; %)**	**Chromosomal Location in the T2T-CHM13 Genome**																							
				1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**X**	**Y**
α-satellite	~171	39	85.67; 2.75	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	▓
β-satellite	68	52	8.61; 0.28	▓	□	░	□	░	□	▓	░	▓	░	□	□	▓	■	▓	░	□	□	▓	░	■	■	□	□
γ-satellite	~217	59	0.65; 0.02	□	□	□	□	□	□	□	▓	░	□	□	▓	░	░	░	□	□	□	□	□	░	░	░	□
HSATI	42 (1A)	22 (1A)	13.39; 0.43 (1A)	□	□	■	■	□	□	□	░	□	□	□	□	■	▓	▓	□	□	□	□	□	▓	▓	□	□
	2420 (1B)	24 (1B)	15.32; 0.49 (1B)	□	□	□	□	□	□	□	□	□	□	□	□	░	░	▓	□	□	□	□	□	░	▓	□	■
HSATII	Variable	37	28.71; 0.92	■	▓	□	□	□	□	▓	□	░	▓	□	□	□	□	□	■	░	□	□	□	□	▓	□	□
HSATIII	Variable	41	69.33; 2.22	▓	□	▓	▓	▓	□	▓	□	■	▓	□	□	■	■	■	░	▓	□	□	▓	■	■	□	□

Black square: >1 Mbp of sequence is present on the indicated chromosome; dark gray square: 100 kbp–1Mbp is present; light gray square: 10–100 kbp is present; white square: <10 kbp is present. Calculations include satellites from chromosomes 1–22, X, and Y in the T2T-CHM13 v2.0 genome.

2. A Complete Census of Centromeric Satellite DNA from One Human Genome

The Telomere-to-Telomere (T2T) Consortium recently resolved the first complete sequence of a human genome (T2T-CHM13) [45], and, in doing so, unveiled the sequence composition of all human centromeres (Figure 2) [44]. There were three advances that made this possible: (1) the use of a complete hydatidiform mole (CHM), where no allelic variation existed to confound assembly of highly repetitive and identical haplotypes; (2) the use of Pacific Biosciences (PacBio) high-fidelity (HiFi) sequence data, which generated a highly accurate sequence backbone of nearly all of the human genome; and (3) the application of ultra-long Oxford Nanopore Technologies (UL-ONT) sequence data, which allowed sequence contigs to be effectively scaffolded. The latter was especially critical to traverse the megabases of repetitive DNA constituting the centromeric regions of the human genome [42–44]. Accompanying these advances in technology were a series of rapid-fire-in-succession genome assembly algorithms, such as HiCanu [46], hifiasm [47], and Verkko [48] that leveraged the unique attributes of the different long-read sequencing technologies or specialized in the characterization and validation of centromeric satellite DNA assemblies [49–51]. It should be noted that subsequent development of methods to accurately phase diploid genomes [52–54] has meant that use of a CHM is no longer necessary, and centromeres from diploid human genomes have now been readily assembled [42,44,48]. However, centromeric regions are still preferentially associated with breaks in standard long-read sequence assemblies, such as in the Human Pangenome Reference [55].

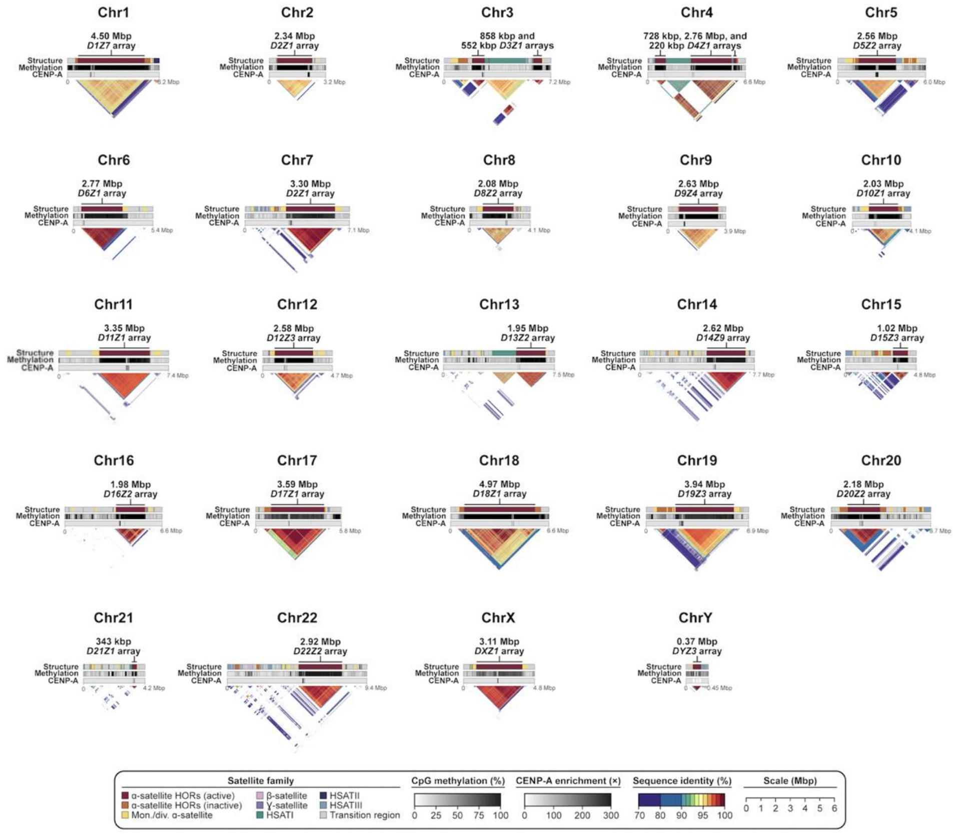

Figure 2. Sequence composition, DNA methylation pattern, and CENP-A chromatin organization of each centromere in the T2T-CHM13 genome. Tracks showing the sequence composition, frequency of methylated CpG dinucleotides, and fold-enrichment of CENP-A ChIP-seq reads over bulk nucleosomal reads (or in the case of chromosome Y, the number of CENP-A CUT and RUN reads) for each centromere in the T2T-CHM13 v2.0 genome. Triangular StainedGlass [56] plots indicate the pairwise sequence identity between 5 kbp segments along each centromeric region and are colored by sequence identity. Warmer colors indicate higher sequence identity, and colder colors indicate lower sequence identity (as indicated in the legend).

Centromeres and their associated pericentromeric DNA are estimated to constitute ~6.2% (189 Mbp) [44] or 7.1% (221.7 Mbp, if the Y chromosome is included) [57] of the human genome and are largely composed of megabases of α-satellite, β-satellite, ɣ-satellite, and three human satellites (HSATI, HSATII, HSATIII) distributed differentially among human chromosomes (Figure 1, Table 1). It should be noted that there is no cytogenetically recognized pericentromere in humans, but the term "pericentromeric DNA" was originally used to describe the five Mbp of DNA extending on either side of the higher-order α-satellite [36]. In humans, pericentromeric DNA contains various classes of inactive α-satellite, almost all other forms of satellites, and large blocks of recent segmental duplication shared among non-homologous chromosomes [36]. It represents the transition region to euchromatin. Later, more nuanced approaches refined this definition to represent the haplotypes flanking the centromere, termed "cenhaps" [58], which tend to evolve as a single chromosomal segment due to infrequent recombination and extensive linkage disequilibrium [44]. In humans, α-satellite, a ~171 bp repeat, is the most abundant, spanning 85.7 Mbp (2.8%) of the human genome and almost exclusively associated with the kinetochore [44,45,57]. Most α-satellite DNA are organized into higher-order arrays consisting of discrete units of monomers repeated in tandem hundreds to thousands of times and flanked on their periphery by divergent HORs and monomeric α-satellite. While most human chromosomes harbor more than one related α-satellite HOR array, only one HOR array is typically associated with the kinetochore, and these are defined as "active" α-satellite HOR arrays. Active α-satellite HOR arrays were first identified in studies of dicentric chromosomes, which have one "active" and one "inactive" centromere. Using immunofluorescence microscopy, these studies revealed that active centromeres were enriched with nucleosomes containing the histone H3 variant CENP-A, while the inactive centromeres were not [59,60]. In general, α-satellite DNA extend contiguously with occasional interruption by mobile element retrotransposition events or blocks of other satellite DNA (for example, the *D3Z1* and *D4Z1* α-satellite HOR arrays on chromosomes 3 and 4, which are disrupted by an array of HSATI repeats). Sequence analysis reveals few examples of inversions within the HOR units, suggesting the orientation of α-satellite is typically maintained [44].

Human centromeric α-satellites are categorized into twenty different suprachromosomal families (SFs; SF1-18, SF01, and SF02), which are groups of α-satellites that are more closely related to each other than to other groups [31,44,61–64]. The sequence identity between α-satellite HORs in an SF is ~85–88%, and between different SFs, it is 50–85% [44]. The three main SFs, SF1-3, represent the "active" (kinetochore-binding) α-satellite HOR arrays on all human autosomes and the X chromosome, and they are composed of either dimeric (SF1 and SF2) or pentameric (SF3) monomer configurations [63]. The SF4 and SF5 families usually flank the active α-satellite HOR arrays and are composed of either purely monomeric repeats (SF4) or a combination of monomeric and divergent HORs that lack a regular repeat structure (SF5) [30,32,63]. Thirteen minor SFs (SF6-18) represent older, more ancient α-satellite monomers that either reside on the extreme edges of centromeric regions or are located far away from the centromere, potential relics of long-defunct ancient centromeres [44]. Finally, SF01 and SF02 are recently defined SF classifications representing mixtures of SFs residing in pericentromeric DNA [44,64,65].

The transition to euchromatin, based on CpG methylation profiling for most chromosomes, is relatively sharp [42–44,66], with the exception of the acrocentric chromosomes (chromosomes 13, 14, 15, 21, and 22), where blocks of α-satellite extend into the short arms of the chromosomes, interdigitating among blocks of segmental duplications (SDs) and rDNA clusters (Figure 2). While SDs and other classes of satellite DNA map peripherally to α-satellite centromere-associated DNA, relatively few functional protein-coding genes have been identified within pericentromeric DNA, consistent with detailed analyses of these regions two decades earlier, which predicted an abundance of unprocessed pseudogenes [36,67]. Of the 676 potential gene annotations, only 23 correspond to validated protein-coding genes, such as *KCNJ17* and *UBBP4*. One of the most remarkable features of pericentromeric DNA is the extent of interchromosomal homology among subsets of human chromosomes (Figure 3). An analysis of various classes of satellite DNA provided compelling evidence that β-satellite, HSATI, and HSATIII share the highest degree of homology among the acrocentric chromosomes (Figure 3) [44]. Similarly, pericentromeric segmental duplications are the largest and most identical among the short arms of chromosomes 13, 14, 15, and 22, with the intervals between the centromeric satellite and secondary constrictions (qh regions) on chromosomes 1, 9, and 16 showing some of the highest degree of interchromosomal homology (Figure 3) [68]. While cause and consequence are difficult to disentangle, it is likely that this homology facilitates association of acrocentric chromosomes to form the nucleolus as well as ectopic exchanges and duplication among specific subsets of nonhomologous chromosomes.

Consistent with original theoretical expectations and subsequent phylogenetic analyses [29,32,34], most centromeric α-satellite HORs are organized into evolutionary layers, with divergent α-satellites residing on the periphery and becoming increasingly more homogenous and displaying high sequence identity within the interior [42,44]. Layers are distinguished from more gradual decay or divergence of HORs by relatively sharp transitions of sequence homology (see numbered arrows for Figure 4a). For some chromosomes, the organization appears highly symmetrical, creating a mirror-like perspective as first noted for chromosome 8 [42] and subsequently observed for chromosomes 17, 18, and 19 [44]. For other centromeres, the α-satellite HOR array is more homogeneously or asymmetrically distributed, although it should be stressed that most of our current understanding has been shaped by the analysis of one human genome. Given the amount of variation observed in sequence and structure among human haplotypes, many more complete centromeres will need to be surveyed. New analytical (e.g., NTPrism [44], HORmon [69], and CentromereArchitect [70]) and visualization (e.g., StainedGlass [56]) tools that were developed to facilitate the identification of these even higher-order patterns within α-satellite and other satellite DNA will be important for future studies of satellite DNA variation and evolution.

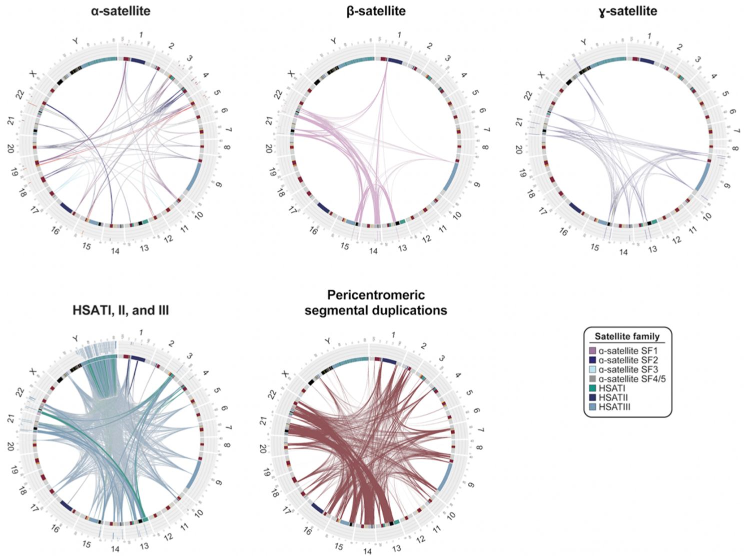

Figure 3. Interchromosomal relationships between centromeric satellite DNA and pericentromeric segmental duplications. Circos plots showing sequence relationships among four different satellite families as well as pericentromeric segmental duplications in the T2T-CHM13 genome. Connecting line widths for satellite families indicate the proportion of 75-mers shared between arrays (i.e., thicker lines indicate greater overall sequence similarity between different arrays of the same family). α-satellite lines are colored by their suprachromosomal family (SF) assignment. Radial bar plots indicate specificity of 75-mers proportionally, with white indicating 75-mers unique to the array, light gray indicating 75-mers shared with other centromeric regions, and black indicating 75-mers shared with regions outside of centromeres. The pericentromeric segmental duplication circos plot shows pairwise alignments between pericentromeric regions that are >1 kbp and >90% identical.

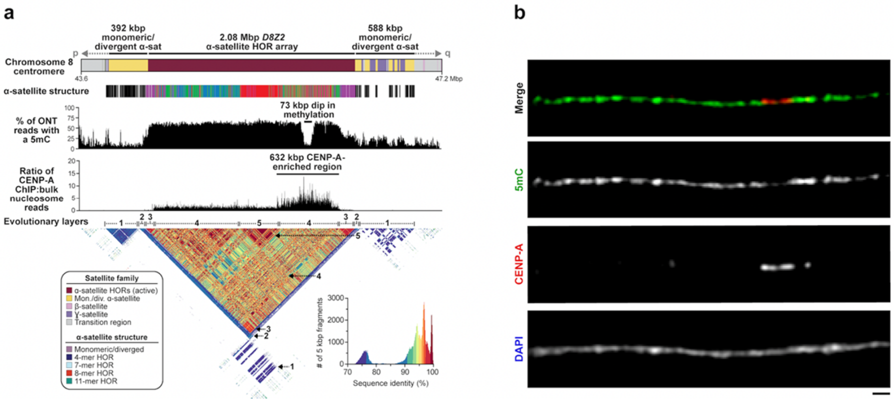

Figure 4. The site of kinetochore attachment within the chromosome 8 centromere. (**a**) Schematic showing the sequence composition, α-satellite structure, CpG methylation frequency, and CENP-A chromatin organization of the chromosome 8 centromere in the T2T-CHM13 genome. The *D8Z2* α-satellite HOR array is 2.08 Mbp long and is generally methylated, except for a 73 kbp region enriched with nucleosomes containing the histone H3 variant CENP-A. CENP-A chromatin resides on structurally diverse α-satellite HORs. (**b**) Representative images of a CHM13 chromatin fiber showing CENP-A enrichment in a hypomethylated region. Scale bar, 1 μm. Adapted from [42].

3. CpG Methylation and the Discovery of the DNA Site of Kinetochore Attachment

In addition to facilitating the assembly of centromeric regions, long-read sequencing also allowed direct detection of CpG-methylated base pairs from native DNA. Initial studies of chromosome 8 and X centromeres both observed a conspicuous region of hypomethylation (approximately 61–73 kbp in length) buried within the hypermethylated active α-satellite HOR array [42,43]. Using chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments, Logsdon was the first to show that this region was enriched for the centromeric histone CENP-A—an observation validated by CENP-A immunostaining on chromatin fibers [42] (Figure 4). In the case of chromosome 8, CENP-A enrichment extended over a broader stretch of 632 kbp, but the peak enrichment centered over the hypomethylated α-satellite HORs. Because CENP-A is a histone H3 variant specifically associated with the centromere [71], these observations suggested that the hypomethylated pocket represents the binding site of the functional kinetochore. The broader CENP-A chromatin peak may represent variability in the position of CENP-A nucleosomes among a population of cells, which has been observed on individual chromatin fibers from native centromeres [72]. Subsequent follow-up experiments, including CUT&RUN, confirmed these general properties on the remainder of the centromeres within the T2T-CHM13 genome [44,66]. The conspicuous hypomethylated region on each centromere was later termed the "centromere dip region" or CDR [44,66]. Genome-wide analyses showed that CDRs were typically constrained to 26–423 kbp in length with CENP-A enrichment extending 190–570 kbp within the active α-satellite HOR array [44]. CDRs mapped only to the active α-satellite HOR arrays, which are typically among the largest and show the highest degree of CpG methylation [66]. In many cases, the CDR and CENP-A enrichment was associated with the evolutionarily youngest and recently expanded α-satellite HOR array, although this was not a universal observation for all centromeres [44]. In the case of chromosome 8, both the CDR and the CENP-A-enriched region mapped to a more diverse set of α-satellite HORs (Figure 4). Similar offsets with respect to the most recently expanded α-satellite HORs were observed for chromosomes 5, 7, and 13. Because the site of kinetochore attachment and hypomethylation are epigenetic properties, caution should be exercised in drawing genetic correlations with the composition of the α-satellite HORs until more genomes and tissues have been examined. Preliminary analyses suggest that the site of kinetochore may, in fact, vary considerably depending on the haplotype in question.

4. Variation in Centromeric Satellite Sequence and Structure

Centromeric satellites are prone to single-nucleotide and structural variation induced by mutational processes such as unequal crossover [29] and gene conversion [73]. These processes can result in rapid changes in satellite sequence composition, repeat structure, and copy number. Early studies using cytogenetic and gel-based techniques revealed that centromeric satellites often undergo repeat amplifications and contractions, which can result in dramatic changes in satellite array size on the order of tens to thousands of kilobases [74,75]. While the detection of large-scale variation in satellite array structure is feasible with cytogenetic and gel-based techniques, discovery of finer-scale variation, such as changes in satellite sequence composition or repeat structure, has been historically difficult to achieve. With complete sequence assemblies of centromeres from multiple human genomes [42,44,76], however, detection of these more fine-scale variants has recently become attainable. A comparison of the chromosome Y *DYZ3* α-satellite HOR array from 21 diverse human genomes using high-quality sequence assemblies, for example, recently enabled the discovery of a 36-mer α-satellite HOR in 52.4% of human haplotypes [76] (Figure 5a). This α-satellite HOR is thought to be an ancestral version of the canonical 34-mer α-satellite HOR, which was born out of repeated deletions of α-satellite monomers at the 22nd monomer position. Similarly, a comparison of the chromosome X *DXZ1* α-satellite HOR array from seven diverse human genomes revealed the presence of duplications spanning hundreds of kilobases in two human genomes (HG01109 and HG03492) as well as the emergence of an α-satellite HOR haplotype predominantly in those with African ancestry [44] (Figure 5b). Finally, pairwise sequence comparisons of the chromosome 8 centromeric region from three human haplotypes revealed gross variation in the sequence and structure of the *D8Z2* α-satellite HOR array, with the pairwise single-nucleotide variant sequence identity dropping down to 99.6% on average, significantly lower than that of the pericentromeric flanking sequences [42] (Figure 5c). As more and more human genomes are sequenced and assembled, the catalog of novel α-satellite sequence and structural variants will become more complete, facilitating more sophisticated models of human centromeric satellite variation. Such models should ultimately distinguish both hypervariable and conserved structural features relevant to chromosome segregation and disease.

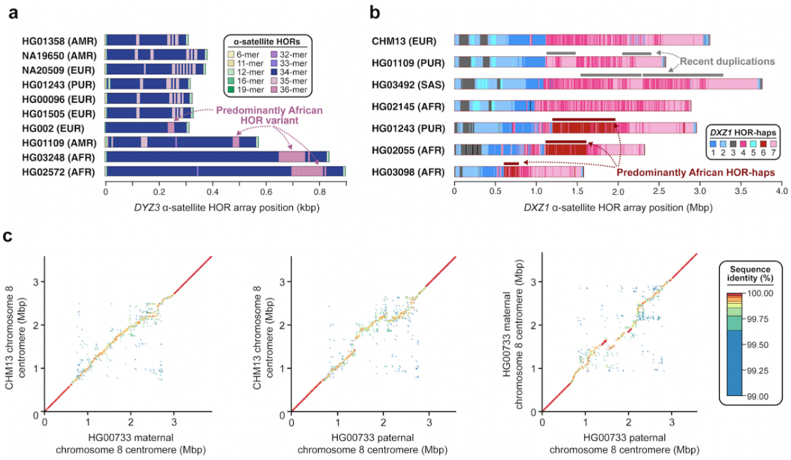

Figure 5. Sequence and structural variation within centromeric α-satellite HOR arrays. (**a**) Plots showing the structure of the chromosome Y *DYZ3* α-satellite HOR array in ten diverse human genomes, highlighting the presence of a 36-mer α-satellite HOR variant in four haplotypes (HG002, HG01109, HG03248, and HG02572) [76]. (**b**) Plots showing the α-satellite HOR haplotypes (HOR-haps) present in the chromosome X *DXZ2* α-satellite HOR array in seven diverse human genomes. Two genomes (HG01109 and HG03492) harbor a haplotype with a recent duplication, while three others (HG01243, HG02055, and HG03098) harbor a haplotype that is especially prevalent among those with African ancestry. Adapted from [44]. (**c**) Plots showing the pairwise sequence identity between chromosome 8 centromeric regions from three human haplotypes (CHM13 and two haplotypes from HG00733). The *D8Z2* α-satellite HOR array shows variation in sequence and structure, while the flanking sequences do not. Adapted from [42].

5. Human Centromere Evolution

Numerous comparative sequence studies between human and nonhuman genomes have shown that centromeric satellites evolve at an accelerated pace due to the action of mutational processes, including concerted gene evolution, saltatory amplification, and unequal crossover [30,32,33]. Estimating the actual increase in mutation rate, however, remains challenging due to the difficulty of sequence alignment [49], even when orthologous centromeres are completely sequenced among closely related species. Comparing the human and chimpanzee chromosome 8 centromeres, for example, the α-satellite HOR array is too divergent to generate a simple pairwise alignment that would permit the computation of a mutation rate over the last six million years since speciation [42] (Figure 6a). Reliable one-to-one alignments, however, have been made spanning the α-satellite monomeric transition regions and, even in these more tractable portions, we find evidence of increased allelic divergence of at least 2.2-fold. Phylogenetic reconstructions focused on the 171 bp α-satellite monomer itself clearly show that the peripheral α-satellite repeats evolve more slowly than the α-satellite HORs, revealing phylogenetic relationships between macaques and humans (~25 million years ago) and defining potential ancestral centromere regions shared among these diverse lineages [32,42]. The analysis also reveals distinct evolutionary trajectories for the emergence of both the dimeric α-satellite (predominant among Old

World monkeys) and the higher-order α-satellite (common to the great ape lineages). The evolutionary turnover of α-satellite HORs is extraordinary as novel α-satellite repeats emerge, amplify, and homogenize through mechanisms such as unequal crossover and gene conversion—a common feature among many species [77]. Different lineages appear to have different characteristics with respect to higher levels of organization; among orangutans, for example, composite HORs have been noted where the HOR layers show relatively little sequence homology among themselves (Figure 6b), in contrast to Old World monkeys, where dimeric repeat structures are distributed among all autosomes [42]. This rapid and stereotypic evolution of centromeres has been described as a potential driving force for speciation, due to the accumulation of sequence differences that result in highly divergent α-satellite HOR sequences and, subsequently, cause incompatibility and reproductive barrier in hybrids between closely related species [78,79]. Even within a species, however, there is extraordinary single-nucleotide and structural variant diversity, possibly as a result of ongoing centromeric meiotic drive to segregate more efficiently [80]. Without this selection, centromeric satellites degrade very quickly, as is evidenced by the structure of the inactivated human chromosome 2 centromere, which reduced from 4.04 Mbp of α-satellite HORs to 41 kbp of divergent α-satellite monomers after the chromosome 2p/2q fusion occurred in the ancestral human lineage (Figure 6c). In this light, it is interesting that sequence comparisons among three human centromere 8 haplotypes show regions of excess allelic variation and structural divergence, although the location and composition of HORs differ among haplotypes [42] (Figure 5c). It is likely that evolutionary reconstructions among different species will be preceded by first reconstructing the dynamic mutational changes that distinguish major haplotypes within species and relating these changes to relocations of kinetochore attachment. Such analyses should, in turn, lead to improved assembly algorithms and improvements in mutational modeling of such complex regions.

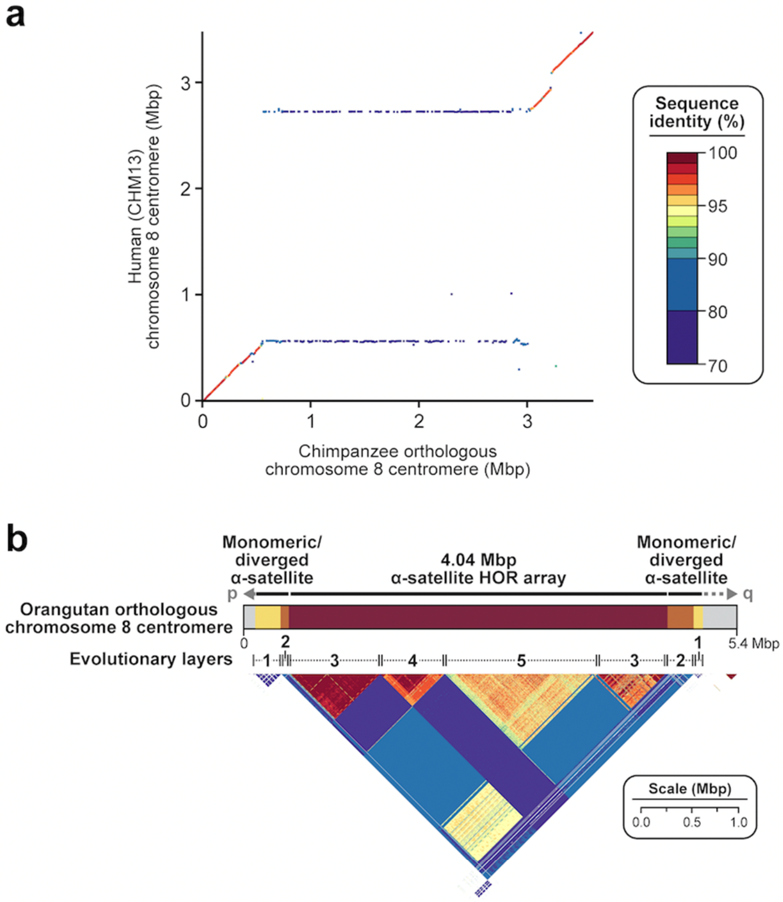

Figure 6. *Cont.*

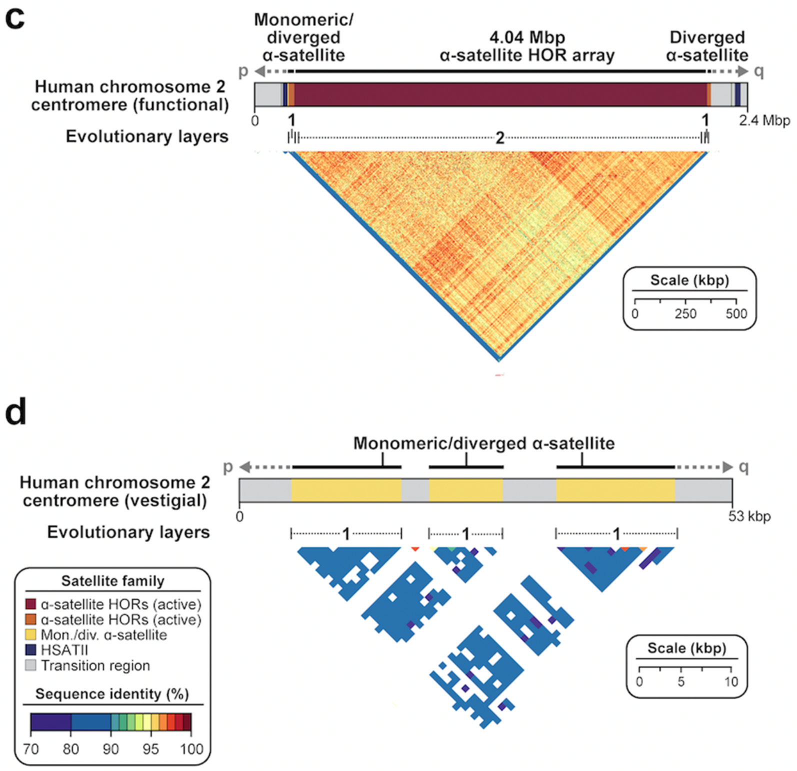

Figure 6. Evolution of centromeric satellite regions. (**a**) Dot matrix plot showing pairwise sequence identity between human (CHM13) and chimpanzee chromosome 8 centromeric regions. While there is >95% sequence identity in the monomeric *α*-satellite sequences and 70–80% sequence identity in the transition regions between monomeric *α*-satellite and *α*-satellite HORs, there is almost no sequence identity shared between sequences in the *α*-satellite HOR array. Despite being similar in size, the two apes show nearly complete turnover of their *α*-satellite HORs, with best matches occurring at the periphery (horizontal lines). (**b**) StainedGlass view [56] showing the organization and pairwise sequence identity among 5 kbp segments across an orangutan orthologous chromosome 8 centromere. Orangutan *α*-satellite HOR arrays are composed of mosaic blocks of *α*-satellite HORs with <95% sequence identity shared between them. Adapted from [42]. (**c**,**d**) StainedGlass view [56] showing the organization and pairwise sequence identity among 5 kbp segments across (**c**) the functional human (CHM13) chromosome 2 centromere and (**d**) the vestigial human (CHM13) chromosome 2 centromere. The vestigial centromere is a remnant of the chromosome 2B centromere present in chimpanzee. It is highly degraded and reduced in size with little obvious *α*-satellite HOR structure, owing to its inactivity.

6. Future Directions

The complete sequence of a human genome has significantly advanced our understanding of the sequence composition, organization, DNA methylation patterns, and chromatin landscape of satellite DNA, but these findings only reflect that of a single human genome. The dynamic nature (Figure 5) and rapid evolution (Figure 6) of human satellite DNA suggest that these sequences are likely to be among the most variable among humans. As such, one representation of satellite sequence and structure is far from sufficient, and many more genomes will need to be sequenced to accurately model genetic variation in these regions. Efforts to assess the variation of satellite sequences among the human population are currently underway, with the Human Pangenome Reference Consortium (HPRC) expected to sequence at least 350 diverse human genomes [81] and the T2T Consortium proposing to finish a subset of these. Given the considerable single-nucleotide and structural variation already observed among a subset of human centromeric haplotypes [42,44,46], it is likely that many more genomes will need to be sequenced and assembled for the genetic diversity of these regions to be sufficiently represented and understood. Accurately representing the complex forms of genetic variation in a graph-based pangenome reference, however, is an unmet challenge requiring significant algorithmic development [82–85]. Other efforts to generate complete and accurate assemblies of nonhuman primates are also ongoing, which will provide the necessary outgroups to reconstruct the evolutionary history of these and other highly dynamic regions. As the cost for long-read sequencing continues to plummet and a USD 1000-long-read-sequenced genome comes within reach [86], we anticipate that the generation of nearly complete, phased human genome assemblies will also become routine. This development is especially important for individuals who have rare or complex diseases with no known genetic or epigenetic cause, who stand to benefit from complete sequence resolution of satellite DNA and other previously unresolved regions of their human genome.

The availability of hundreds of thousands of genomes from both healthy and diseased individuals will also advance our understanding of centromere biology. In particular, it will allow one to delineate the genetic relationship between satellite DNA variation and the site of kinetochore attachment, if one exists. It is possible and even likely that variation in the α-satellite HOR array structure and/or chromatin landscape affects the accuracy of chromosome segregation during cell division, and such differences may contribute to infertility, trisomy disorders, and aneuploidy associated with cancer. Detection of pathogenic variants within these regions, whether genetic or epigenetic, will also benefit from combining long-read sequencing data with electronic health records from thousands of individuals to discover significant associations with disease. Federally funded efforts, such as the NIH *All of Us* Research Program, which aims to make de-identified genomic sequencing data and medical records available, has recently embarked on a long-read sequencing initiative to generate data from more than 10,000 genomes [87]. Such efforts promise to accelerate research into these more complex regions of the genome and will serve as a model for other biobank efforts in the future. Critical to this is the continued commitment that genomic data, once de-identified, should become publicly available in order to advance both basic and translational biomedical research. This approach, which was initiated and widely accepted in the early days of the Human Genome Project, has benefitted the greater scientific research communities and will continue to accelerate and democratize genomic research as we begin to access some of the most genetically complex and dynamic regions of our genomes.

Author Contributions: G.A.L. and E.E.E. conceived the content and organization of the manuscript and wrote the manuscript. G.A.L. illustrated the figures. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported, in part, by U.S. National Institutes of Health (NIH) grants K99GM147352 to G.A.L. and R01HG010169 to E.E.E., E.E.E. is an investigator of the Howard Hughes Medical Institute.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The sequence of the T2T-CHM13 v2.0 genome is available at GenBank accession GCA_009914755.4, and the sequence of the human and nonhuman primate chromosome 8 centromere assemblies are available at GenBank accessions OM502025-OM502028. The Human Pangenome Reference Consortium (HPRC) whole-genome long-read sequence assemblies are available at https://github.com/human-pangenomics/HPP_Year1_Assemblies (accessed on 14 July 2021), and the Human Genome Structural Variation Consortium (HGSVC) whole-genome long-read sequence assemblies are available at http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HGSVC2/release/v1.0/assemblies/ (accessed on 5 June 2021). The data and code used to generate the circos plots are available at the following GitHub repositories: https://github.com/msauria/T2T_Kmer_Analysis (for satellites; accessed on 9 December 2022) and https://github.com/mrvollger/Data-Analysis-for-SDs-in-T2T-CHM13 (for segmental duplications; accessed on 9 December 2022).

Acknowledgments: The authors thank M.E.G.S. (JHU) and M.R.V. (UW) for assistance with analysis and T.B. for assistance in editing this manuscript.

Conflicts of Interest: E.E.E. is a scientific advisory board (SAB) member of Variant Bio, Inc.

References

1. Corneo, G.; Ginelli, E.; Polli, E. A Satellite DNA Isolated from Human Tissues. *J. Mol. Biol.* **1967**, *23*, 619–622. [CrossRef] [PubMed]
2. Corneo, G.; Ginelli, E.; Polli, E. Presence of a Satellite DNA in Normal and Leukemic Human Tissues. *AHA* **1968**, *39*, 75–84. [CrossRef] [PubMed]
3. Corneo, G.; Ginelli, E.; Polli, E. Repeated Sequences in Human DNA. *J. Mol. Biol.* **1970**, *48*, 319–327. [CrossRef] [PubMed]
4. Miga, K.H.; Newton, Y.; Jain, M.; Altemose, N.; Willard, H.F.; Kent, W.J. Centromere Reference Models for Human Chromosomes X and Y Satellite Arrays. *Genome Res.* **2014**, *24*, 697–707. [CrossRef]
5. Jones, K.W.; Purdom, I.F.; Prosser, J.; Corneo, G. The Chromosomal Localisation of Human Satellite DNA I. *Chromosoma* **1974**, *49*, 161–171. [CrossRef]
6. Jones, K.W.; Prosser, J.; Corneo, G.; Ginelli, E. The Chromosomal Location of Human Satellite DNA 3. *Chromosoma* **1973**, *42*, 445–451. [CrossRef]
7. Gosden, J.R.; Mitchell, A.R.; Buckland, R.A.; Clayton, R.P.; Evans, H.J. The Location of Four Human Satellite DNAs on Human Chromosomes. *Exp. Cell Res.* **1975**, *92*, 148–158. [CrossRef]
8. Prosser, J.; Reisner, A.H.; Bradley, M.L.; Ho, K.; Vincent, P.C. Buoyant Density and Hybridization Analysis of Human DNA Sequences, Including Three Satellite DNAs. *Biochim. Biophys. Acta (BBA)—Nucleic Acids Protein Synth.* **1981**, *656*, 93–102. [CrossRef]
9. Prosser, J.; Frommer, M.; Paul, C.; Vincent, P.C. Sequence Relationships of Three Human Satellite DNAs. *J. Mol. Biol.* **1986**, *187*, 145–155. [CrossRef]
10. Corneo, G.; Zardi, L.; Polli, E. Elution of Human Satellite DNAs on a Methylated Albumin Kieselguhr Chromatographic Column:Isolation of Satellite DNA IV. *Biochim. Biophys. Acta (BBA)—Nucleic Acids Protein Synth.* **1972**, *269*, 201–204. [CrossRef]
11. Frommer, M.; Prosser, J.; Tkachuk, D.; Reisner, A.H.; Vincent, P.C. Simple Repeated Sequences in Human Satellite DNA. *Nucleic Acids Res.* **1982**, *10*, 547–563. [CrossRef] [PubMed]
12. Manuelidis, L. Complex and Simple Sequences in Human Repeated DNAs. *Chromosoma* **1978**, *66*, 23–32. [CrossRef]
13. Willard, H.F. Chromosome-Specific Organization of Human Alpha Satellite DNA. *Am. J. Hum. Genet.* **1985**, *37*, 524. [PubMed]
14. Willard, H.F.; Wevrick, R.; Warburton, P.E. Human Centromere Structure: Organization and Potential Role of Alpha Satellite DNA. *Prog. Clin. Biol. Res.* **1989**, *318*, 9–18. [PubMed]
15. Wu, J.C.; Manuelidis, L. Sequence Definition and Organization of a Human Repeated DNA. *J. Mol. Biol.* **1980**, *142*, 363–386. [CrossRef] [PubMed]
16. Cooke, H.J.; Hindley, J. Cloning of Human Satellite III DNA: Different Components Are on Different Chromosomes. *Nucleic Acids Res.* **1979**, *6*, 3177–3197. [CrossRef] [PubMed]
17. Kalitsis, P.; Earle, E.; Vissel, B.; Shaffer, L.G.; Choo, K.H. A Chromosome 13-Specific Human Satellite I DNA Subfamily with Minor Presence on Chromosome 21: Further Studies on Robertsonian Translocations. *Genomics* **1993**, *16*, 104–112. [CrossRef]
18. Meyne, J.; Goodwin, E.H.; Moyzis, R.K. Chromosome Localization and Orientation of the Simple Sequence Repeat of Human Satellite I DNA. *Chromosoma* **1994**, *103*, 99–103. [CrossRef]
19. Cooke, H.J.; Schmidtke, J.; Gosden, J.R. Characterisation of a Human Y Chromosome Repeated Sequence and Related Sequences in Higher Primates. *Chromosoma* **1982**, *87*, 491–502. [CrossRef]
20. Nakahori, Y.; Mitani, K.; Yamada, M.; Nakagome, Y. A Human Y-Chromosome Specific Repeated DNA Family (DYZ1) Consists of a Tandem Array of Pentanucleotides. *Nucleic Acids Res.* **1986**, *14*, 7569–7580. [CrossRef]
21. Higgins, M.J.; Wang, H.; Shtromas, I.; Haliotis, T.; Roder, J.C.; Holden, J.J.A.; White, B.N. Organization of a Repetitive Human 1.8 Kb KpnI Sequence Localized in the Heterochromatin of Chromosome 15. *Chromosoma* **1985**, *93*, 77–86. [CrossRef] [PubMed]
22. Choo, K.H.; Earle, E.; McQuillan, C. A Homologous Subfamily of Satellite III DNA on Human Chromosomes 14 and 22. *Nucleic Acids Res.* **1990**, *18*, 5641–5648. [CrossRef] [PubMed]
23. Tagarro, I.; Wiegant, J.; Raap, A.K.; González-Aguilera, J.J.; Fernández-Peralta, A.M. Assignment of Human Satellite 1 DNA as Revealed by Fluorescent in Situ Hybridization with Oligonucleotides. *Hum. Genet.* **1994**, *93*, 125–128. [CrossRef] [PubMed]
24. Lin, C.C.; Sasi, R.; Lee, C.; Fan, Y.S.; Court, D. Isolation and Identification of a Novel Tandemly Repeated DNA Sequence in the Centromeric Region of Human Chromosome 8. *Chromosoma* **1993**, *102*, 333–339. [CrossRef]
25. Greig, G.M.; Willard, H.F. Beta Satellite DNA: Characterization and Localization of Two Subfamilies from the Distal and Proximal Short Arms of the Human Acrocentric Chromosomes. *Genomics* **1992**, *12*, 573–580. [CrossRef]
26. Moyzis, R.K.; Albright, K.L.; Bartholdi, M.F.; Cram, L.S.; Deaven, L.L.; Hildebrand, C.E.; Joste, N.E.; Longmire, J.L.; Meyne, J.; Schwarzacher-Robinson, T. Human Chromosome-Specific Repetitive DNA Sequences: Novel Markers for Genetic Analysis. *Chromosoma* **1987**, *95*, 375–386. [CrossRef]
27. Tagarro, I.; Fernández-Peralta, A.M.; González-Aguilera, J.J. Chromosomal Localization of Human Satellites 2 and 3 by a FISH Method Using Oligonucleotides as Probes. *Hum. Genet.* **1994**, *93*, 383–388. [CrossRef]

28. Schwarzacher-Robinson, T.; Cram, L.S.; Meyne, J.; Moyzis, R.K. Characterization of Human Heterochromatin by in Situ Hybridization with Satellite DNA Clones. *Cytogenet. Cell Genet.* **1988**, *47*, 192–196. [CrossRef]
29. Smith, G.P. Evolution of Repeated DNA Sequences by Unequal Crossover. *Science* **1976**, *191*, 528–535. [CrossRef] [PubMed]
30. Alexandrov, I.; Kazakov, A.; Tumeneva, I.; Shepelev, V.; Yurov, Y. Alpha-Satellite DNA of Primates: Old and New Families. *Chromosoma* **2001**, *110*, 253–266. [CrossRef]
31. Alexandrov, I.A.; Mitkevich, S.P.; Yurov, Y.B. The Phylogeny of Human Chromosome Specific Alpha Satellites. *Chromosoma* **1988**, *96*, 443–453. [CrossRef]
32. Shepelev, V.A.; Alexandrov, A.A.; Yurov, Y.B.; Alexandrov, I.A. The Evolutionary Origin of Man Can Be Traced in the Layers of Defunct Ancestral Alpha Satellites Flanking the Active Centromeres of Human Chromosomes. *PLoS Genet.* **2009**, *5*, e1000641. [CrossRef]
33. Alkan, C.; Eichler, E.E.; Bailey, J.A.; Sahinalp, S.C.; Tüzün, E. The Role of Unequal Crossover in Alpha-Satellite DNA Evolution: A Computational Analysis. *J. Comput. Biol.* **2004**, *11*, 933–944. [CrossRef] [PubMed]
34. Alkan, C.; Ventura, M.; Archidiacono, N.; Rocchi, M.; Sahinalp, S.C.; Eichler, E.E. Organization and Evolution of Primate Centromeric DNA from Whole-Genome Shotgun Sequence Data. *PLoS Comput. Biol.* **2007**, *3*, 1807–1818. [CrossRef] [PubMed]
35. Koga, A.; Hirai, Y.; Terada, S.; Jahan, I.; Baicharoen, S.; Arsaithamkul, V.; Hirai, H. Evolutionary Origin of Higher-Order Repeat Structure in Alpha-Satellite DNA of Primate Centromeres. *DNA Res.* **2014**, *21*, 407–415. [CrossRef] [PubMed]
36. She, X.; Horvath, J.E.; Jiang, Z.; Liu, G.; Furey, T.S.; Christ, L.; Clark, R.; Graves, T.; Gulden, C.L.; Alkan, C.; et al. The Structure and Evolution of Centromeric Transition Regions within the Human Genome. *Nature* **2004**, *430*, 857–864. [CrossRef]
37. Warburton, P.E.; Haaf, T.; Gosden, J.; Lawson, D.; Willard, H.F. Characterization of a Chromosome-Specific Chimpanzee Alpha Satellite Subset: Evolutionary Relationship to Subsets on Human Chromosomes. *Genomics* **1996**, *33*, 220–228. [CrossRef]
38. Manuelidis, L.; Wu, J.C. Homology between Human and Simian Repeated DNA. *Nature* **1978**, *276*, 92–94. [CrossRef]
39. Flemming, W. Beiträge Zur Kenntnis Der Zelle Und Ihrer Lebenserscheinungen. *Arch. Mikrosk. Anat.* **1880**, *18*, 151–259. [CrossRef]
40. Moroi, Y.; Peebles, C.; Fritzler, M.J.; Steigerwald, J.; Tan, E.M. Autoantibody to Centromere (Kinetochore) in Scleroderma Sera. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 1627–1631. [CrossRef]
41. Henikoff, J.G.; Thakur, J.; Kasinathan, S.; Henikoff, S. A Unique Chromatin Complex Occupies Young α-Satellite Arrays of Human Centromeres. *Sci. Adv.* **2015**, *1*, e1400234. [CrossRef] [PubMed]
42. Logsdon, G.A.; Vollger, M.R.; Hsieh, P.; Mao, Y.; Liskovykh, M.A.; Koren, S.; Nurk, S.; Mercuri, L.; Dishuck, P.C.; Rhie, A.; et al. The Structure, Function and Evolution of a Complete Human Chromosome 8. *Nature* **2021**, *593*, 101–107. [CrossRef] [PubMed]
43. Miga, K.H.; Koren, S.; Rhie, A.; Vollger, M.R.; Gershman, A.; Bzikadze, A.; Brooks, S.; Howe, E.; Porubsky, D.; Logsdon, G.A.; et al. Telomere-to-Telomere Assembly of a Complete Human X Chromosome. *Nature* **2020**, *585*, 79–84. [CrossRef] [PubMed]
44. Altemose, N.; Logsdon, G.A.; Bzikadze, A.V.; Sidhwani, P.; Langley, S.A.; Caldas, G.V.; Hoyt, S.J.; Uralsky, L.; Ryabov, F.D.; Shew, C.J.; et al. Complete Genomic and Epigenetic Maps of Human Centromeres. *Science* **2022**, *376*, eabl4178. [CrossRef]
45. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The Complete Sequence of a Human Genome. *Science* **2022**, *376*, 44–53. [CrossRef]
46. Nurk, S.; Walenz, B.P.; Rhie, A.; Vollger, M.R.; Logsdon, G.A.; Grothe, R.; Miga, K.H.; Eichler, E.E.; Phillippy, A.M.; Koren, S. HiCanu: Accurate Assembly of Segmental Duplications, Satellites, and Allelic Variants from High-Fidelity Long Reads. *Genome Res.* **2020**, *30*, 1291–1305. [CrossRef] [PubMed]
47. Cheng, H.; Concepcion, G.T.; Feng, X.; Zhang, H.; Li, H. Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm. *Nat. Methods* **2021**, *18*, 170–175. [CrossRef]
48. Rautiainen, M.; Nurk, S.; Walenz, B.P.; Logsdon, G.A.; Porubsky, D.; Rhie, A.; Eichler, E.E.; Phillippy, A.M.; Koren, S. Verkko: Telomere-to-Telomere Assembly of Diploid Chromosomes. *bioRxiv* **2022**. [CrossRef]
49. Bzikadze, A.V.; Pevzner, P.A. TandemAligner: A New Parameter-Free Framework for Fast Sequence Alignment. *bioRxiv* **2022**. [CrossRef]
50. Mikheenko, A.; Bzikadze, A.V.; Gurevich, A.; Miga, K.H.; Pevzner, P.A. TandemTools: Mapping Long Reads and Assessing/Improving Assembly Quality in Extra-Long Tandem Repeats. *Bioinformatics* **2020**, *36*, i75–i83. [CrossRef]
51. Bzikadze, A.V.; Mikheenko, A.; Pevzner, P.A. Fast and Accurate Mapping of Long Reads to Complete Genome Assemblies with VerityMap. *Genome Res.* **2022**, *32*, 2107–2118. [CrossRef] [PubMed]
52. Hanlon, V.C.T.; Porubsky, D.; Lansdorp, P.M. Chromosome-Length Haplotypes with StrandPhaseR and Strand-Seq. *Methods Mol. Biol.* **2023**, *2590*, 183–200. [CrossRef] [PubMed]
53. Kronenberg, Z.N.; Rhie, A.; Koren, S.; Concepcion, G.T.; Peluso, P.; Munson, K.M.; Porubsky, D.; Kuhn, K.; Mueller, K.A.; Low, W.Y.; et al. Extended Haplotype-Phasing of Long-Read de Novo Genome Assemblies Using Hi-C. *Nat. Commun.* **2021**, *12*, 1935. [CrossRef] [PubMed]
54. Chin, C.-S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A.; et al. Phased Diploid Genome Assembly with Single-Molecule Real-Time Sequencing. *Nat. Methods* **2016**, *13*, 1050–1054. [CrossRef] [PubMed]
55. Liao, W.-W.; Asri, M.; Ebler, J.; Doerr, D.; Haukness, M.; Hickey, G.; Lu, S.; Lucas, J.K.; Monlong, J.; Abel, H.J.; et al. A Draft Human Pangenome Reference. *bioRxiv* **2022**. [CrossRef]
56. Vollger, M.R.; Kerpedjiev, P.; Phillippy, A.M.; Eichler, E.E. StainedGlass: Interactive Visualization of Massive Tandem Repeat Structures with Identity Heatmaps. *Bioinformatics* **2022**, *38*, 2049–2051. [CrossRef]

57. Rhie, A.; Nurk, S.; Cechova, M.; Hoyt, S.J.; Taylor, D.J.; Altemose, N.; Hook, P.W.; Koren, S.; Rautiainen, M.; Alexandrov, I.A.; et al. The Complete Sequence of a Human Y Chromosome. *bioRxiv* **2022**. [CrossRef]
58. Langley, S.A.; Miga, K.H.; Karpen, G.H.; Langley, C.H. Haplotypes Spanning Centromeric Regions Reveal Persistence of Large Blocks of Archaic DNA. *eLife* **2019**, *8*, e42989. [CrossRef]
59. Earnshaw, W.C.; Migeon, B.R. Three Related Centromere Proteins are Absent from the Inactive Centromere of a Stable Isodicentric Chromosome. *Chromosoma* **1985**, *92*, 290–296. [CrossRef]
60. Warburton, P.E.; Cooke, C.A.; Bourassa, S.; Vafa, O.; Sullivan, B.A.; Stetten, G.; Gimelli, G.; Warburton, D.; Tyler-Smith, C.; Sullivan, K.F.; et al. Immunolocalization of CENP-A Suggests a Distinct Nucleosome Structure at the Inner Kinetochore Plate of Active Centromeres. *Curr. Biol.* **1997**, *7*, 901–904. [CrossRef] [PubMed]
61. Alexandrov, I.A.; Mashkova, T.D.; Akopian, T.A.; Medvedev, L.I.; Kisselev, L.L.; Mitkevich, S.P.; Yurov, Y.B. Chromosome-Specific Alpha Satellites: Two Distinct Families on Human Chromosome 18. *Genomics* **1991**, *11*, 15–23. [CrossRef]
62. Alexandrov, I.A.; Mashkova, T.D.; Romanova, L.Y.; Yurov, Y.B.; Kisselev, L.L. Segment Substitutions in Alpha Satellite DNA. Unusual Structure of Human Chromosome 3-Specific Alpha Satellite Repeat Unit. *J. Mol. Biol.* **1993**, *231*, 516–520. [CrossRef] [PubMed]
63. Alexandrov, I.A.; Medvedev, L.I.; Mashkova, T.D.; Kisselev, L.L.; Romanova, L.Y.; Yurov, Y.B. Definition of a New Alpha Satellite Suprachromosomal Family Characterized by Monomeric Organization. *Nucleic Acids Res.* **1993**, *21*, 2209–2215. [CrossRef] [PubMed]
64. Uralsky, L.I.; Shepelev, V.A.; Alexandrov, A.A.; Yurov, Y.B.; Rogaev, E.I.; Alexandrov, I.A. Classification and Monomer-by-Monomer Annotation Dataset of Suprachromosomal Family 1 Alpha Satellite Higher-Order Repeats in Hg38 Human Genome Assembly. *Data Brief* **2019**, *24*, 103708. [CrossRef]
65. Shepelev, V.A.; Uralsky, L.I.; Alexandrov, A.A.; Yurov, Y.B.; Rogaev, E.I.; Alexandrov, I.A. Annotation of Suprachromosomal Families Reveals Uncommon Types of Alpha Satellite Organization in Pericentromeric Regions of Hg38 Human Genome Assembly. *Genom. Data* **2015**, *5*, 139–146. [CrossRef] [PubMed]
66. Gershman, A.; Sauria, M.E.G.; Guitart, X.; Vollger, M.R.; Hook, P.W.; Hoyt, S.J.; Jain, M.; Shumate, A.; Razaghi, R.; Koren, S.; et al. Epigenetic Patterns in a Complete Human Genome. *Science* **2022**, *376*, eabj5089. [CrossRef] [PubMed]
67. Horvath, J.E.; Bailey, J.A.; Locke, D.P.; Eichler, E.E. Lessons from the Human Genome: Transitions between Euchromatin and Heterochromatin. *Hum. Mol. Genet.* **2001**, *10*, 2215–2223. [CrossRef]
68. Vollger, M.R.; Guitart, X.; Dishuck, P.C.; Mercuri, L.; Harvey, W.T.; Gershman, A.; Diekhans, M.; Sulovari, A.; Munson, K.M.; Lewis, A.P.; et al. Segmental Duplications and Their Variation in a Complete Human Genome. *Science* **2022**, *376*, eabj6965. [CrossRef]
69. Kunyavskaya, O.; Dvorkina, T.; Bzikadze, A.V.; Alexandrov, I.A.; Pevzner, P.A. Automated Annotation of Human Centromeres with HORmon. *Genome Res.* **2022**, *32*, 1137–1151. [CrossRef]
70. Dvorkina, T.; Kunyavskaya, O.; Bzikadze, A.V.; Alexandrov, I.; Pevzner, P.A. CentromereArchitect: Inference and Analysis of the Architecture of Centromeres. *Bioinformatics* **2021**, *37*, i196–i204. [CrossRef]
71. Black, B.E.; Bassett, E.A. The Histone Variant CENP-A and Centromere Specification. *Curr. Opin. Cell Biol.* **2008**, *20*, 91–100. [CrossRef] [PubMed]
72. Altemose, N.; Maslan, A.; Smith, O.K.; Sundararajan, K.; Brown, R.R.; Mishra, R.; Detweiler, A.M.; Neff, N.; Miga, K.H.; Straight, A.F.; et al. DiMeLo-seq: A Long-Read, Single-Molecule Method for Mapping Protein–DNA Interactions Genome Wide. *Nat. Methods* **2022**, *19*, 711–723. [CrossRef] [PubMed]
73. Schindelhauer, D.; Schwarz, T. Evidence for a Fast, Intrachromosomal Conversion Mechanism from Mapping of Nucleotide Variants within a Homogeneous α-Satellite DNA Array. *Genome Res.* **2002**, *12*, 1815–1826. [CrossRef] [PubMed]
74. Waye, J.S.; Willard, H.F. Molecular Analysis of a Deletion Polymorphism in Alpha Satellite of Human Chromosome 17: Evidence for Homologous Unequal Crossing-over and Subsequent Fixation. *Nucleic Acids Res.* **1986**, *14*, 6915–6927. [CrossRef] [PubMed]
75. Waye, J.S.; Willard, H.F. Structure, Organization, and Sequence of Alpha Satellite DNA from Human Chromosome 17: Evidence for Evolution by Unequal Crossing-over and an Ancestral Pentamer Repeat Shared with the Human X Chromosome. *Mol. Cell. Biol.* **1986**, *6*, 3156–3165. [PubMed]
76. Hallast, P.; Ebert, P.; Loftus, M.; Yilmaz, F.; Audano, P.A.; Logsdon, G.A.; Bonder, M.J.; Zhou, W.; Höps, W.; Kim, K.; et al. Assembly of 43 Diverse Human Y Chromosomes Reveals Extensive Complexity and Variation. *bioRxiv* **2022**. [CrossRef]
77. Plohl, M.; Meštrović, N.; Mravinac, B. Centromere Identity from the DNA Point of View. *Chromosoma* **2014**, *123*, 313–325. [CrossRef]
78. Henikoff, S.; Ahmad, K.; Malik, H.S. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science* **2001**, *293*, 1098–1102. [CrossRef] [PubMed]
79. Malik, H.S.; Henikoff, S. Adaptive Evolution of Cid, a Centromere-Specific Histone in Drosophila. *Genetics* **2001**, *157*, 1293–1298. [CrossRef] [PubMed]
80. Suzuki, Y.; Myers, E.W.; Morishita, S. Rapid and Ongoing Evolution of Repetitive Sequence Structures in Human Centromeres. *Sci. Adv.* **2020**, *6*, eabd9230. [CrossRef]
81. Wang, T.; Antonacci-Fulton, L.; Howe, K.; Lawson, H.A.; Lucas, J.K.; Phillippy, A.M.; Popejoy, A.B.; Asri, M.; Carson, C.; Chaisson, M.J.P.; et al. The Human Pangenome Project: A Global Resource to Map Genomic Diversity. *Nature* **2022**, *604*, 437–446. [CrossRef] [PubMed]

82. Garrison, E.; Guarracino, A. Unbiased Pangenome Graphs. *Bioinformatics* **2022**, btac743. [CrossRef] [PubMed]
83. Garrison, E.; Sirén, J.; Novak, A.M.; Hickey, G.; Eizenga, J.M.; Dawson, E.T.; Jones, W.; Garg, S.; Markello, C.; Lin, M.F.; et al. Variation Graph Toolkit Improves Read Mapping by Representing Genetic Variation in the Reference. *Nat. Biotechnol.* **2018**, *36*, 875–879. [CrossRef] [PubMed]
84. Eizenga, J.M.; Novak, A.M.; Kobayashi, E.; Villani, F.; Cisar, C.; Heumos, S.; Hickey, G.; Colonna, V.; Paten, B.; Garrison, E. Efficient Dynamic Variation Graphs. *Bioinformatics* **2021**, *36*, 5139–5144. [CrossRef] [PubMed]
85. Li, H.; Feng, X.; Chu, C. The Design and Construction of Reference Pangenome Graphs with Minigraph. *Genome Biol.* **2020**, *21*, 265. [CrossRef] [PubMed]
86. PacBio Announces Revio, a Revolutionary New Long Read Sequencing System Designed to Provide 15 Times More HiFi Data and Human Genomes at Scale for Under $1000. Available online: https://www.pacb.com/press_releases/pacbio-announces-revio-a-revolutionary-new-long-read-sequencing-system-designed-to-provide-15-times-more-hifi-data-and-human-genomes-at-scale-for-under-1000/ (accessed on 15 December 2022).
87. NIH Funds New All of Us Research Program Genome Center to Test Advanced Sequencing Tools. Available online: https://allofus.nih.gov/news-events/announcements/nih-funds-new-all-us-research-program-genome-center-test-advanced-sequencing-tools (accessed on 15 December 2022).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.