

Review

# Dark Matter of Primate Genomes: Satellite DNA Repeats and Their Evolutionary Dynamics

Syed Farhan Ahmad <sup>1,2</sup>, Worapong Singchat <sup>1,2</sup>, Maryam Jehangir <sup>1,3</sup>, Aorarat Suntronpong <sup>1,2</sup>, Thitipong Panthum <sup>1,2</sup>, Suchinda Malaivijitnond <sup>4,5</sup> and Kornorn Srikulnath <sup>1,2,4,6,7,\*</sup> 

<sup>1</sup> Laboratory of Animal Cytogenetics and Comparative Genomics (ACCG), Department of Genetics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand; farhan.phd.unesp@gmail.com (S.F.A.); worapong.si@ku.th (W.S.); maryam.bioinfo.unesp@gmail.com (M.J.); aorarat.sun@ku.th (A.S.); thitipong.pa@ku.th (T.P.)

<sup>2</sup> Special Research Unit for Wildlife Genomics (SRUWG), Department of Forest Biology, Faculty of Forestry, Kasetsart University, Bangkok 10900, Thailand

<sup>3</sup> Department of Structural and Functional Biology, Institute of Bioscience at Botucatu, São Paulo State University (UNESP), Botucatu, São Paulo 18618-689, Brazil

<sup>4</sup> National Primate Research Center of Thailand, Chulalongkorn University, Saraburi 18110, Thailand; suchinda.m@chula.ac.th

<sup>5</sup> Department of Biology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

<sup>6</sup> Center of Excellence on Agricultural Biotechnology (AG-BIO/PERDO-CHE), Bangkok 10900, Thailand

<sup>7</sup> Omics Center for Agriculture, Bioresources, Food and Health, Kasetsart University (OmiKU), Bangkok 10900, Thailand

\* Correspondence: kornorn.s@ku.ac.th

Received: 27 October 2020; Accepted: 16 December 2020; Published: 18 December 2020



**Abstract:** A substantial portion of the primate genome is composed of non-coding regions, so-called “dark matter”, which includes an abundance of tandemly repeated sequences called satellite DNA. Collectively known as the satellitome, this genomic component offers exciting evolutionary insights into aspects of primate genome biology that raise new questions and challenge existing paradigms. A complete human reference genome was recently reported with telomere-to-telomere human X chromosome assembly that resolved hundreds of dark regions, encompassing a 3.1 Mb centromeric satellite array that had not been identified previously. With the recent exponential increase in the availability of primate genomes, and the development of modern genomic and bioinformatics tools, extensive growth in our knowledge concerning the structure, function, and evolution of satellite elements is expected. The current state of knowledge on this topic is summarized, highlighting various types of primate-specific satellite repeats to compare their proportions across diverse lineages. Inter- and intraspecific variation of satellite repeats in the primate genome are reviewed. The functional significance of these sequences is discussed by describing how the transcriptional activity of satellite repeats can affect gene expression during different cellular processes. Sex-linked satellites are outlined, together with their respective genomic organization. Mechanisms are proposed whereby satellite repeats might have emerged as novel sequences during different evolutionary phases. Finally, the main challenges that hinder the detection of satellite DNA are outlined and an overview of the latest methodologies to address technological limitations is presented.

**Keywords:** non-human primates; alpha satellite; tandem repeats; heterochromatin; centromere; evolution

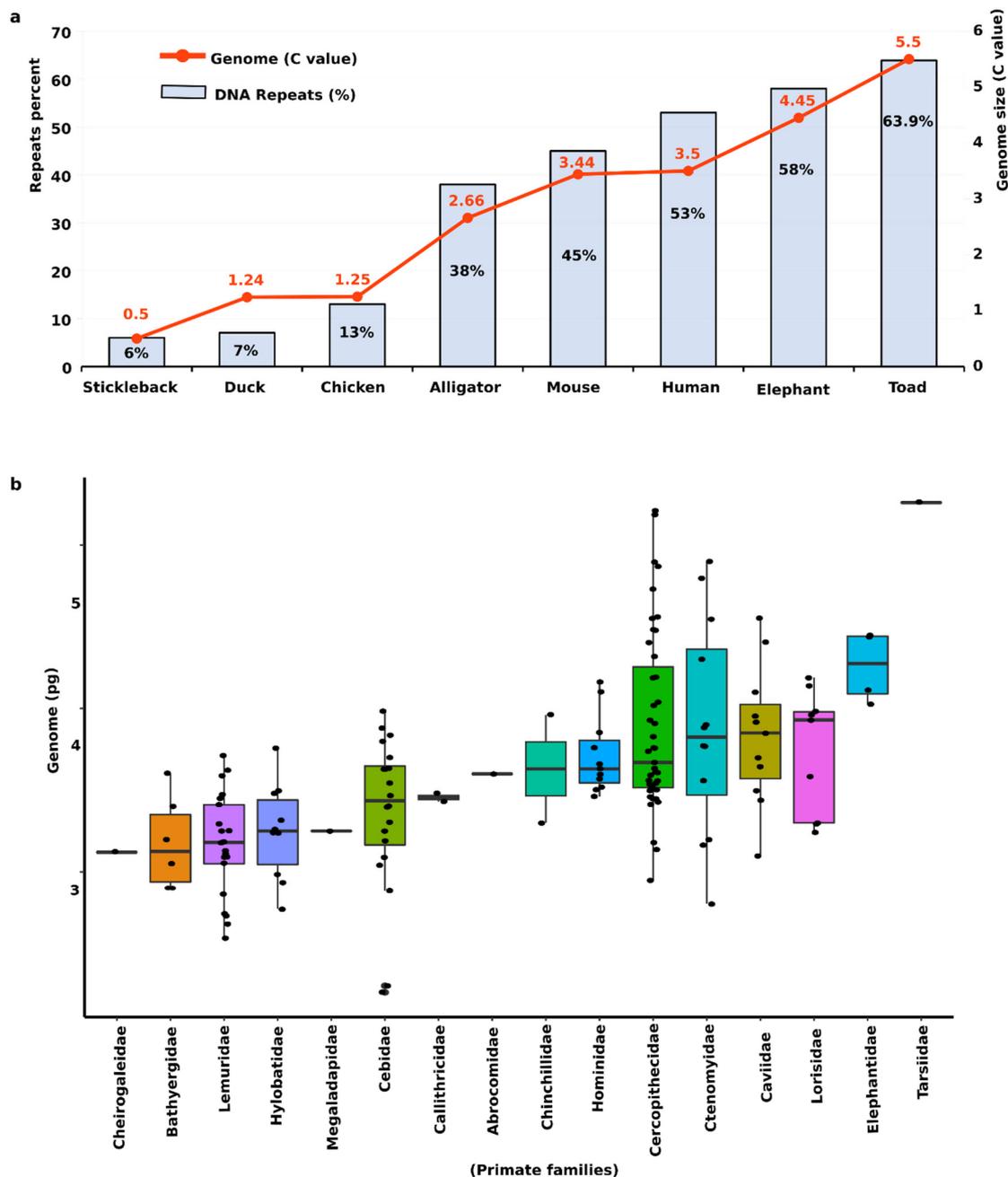
## 1. Introduction

The latest advances in genome sequencing technologies and an increase in the number of available genomes have presented novel opportunities for comparative and evolutionary genomics research.

Increased knowledge regarding the patterns of primate genome contents and dynamics, such as great apes and macaques, has generated critical information regarding the evolutionary origin of the human genome and related biomedical predictions [1]. Assemblies of non-human primate (NHP) genomes have provided excellent resources to study genetic variations and similarities of model species used for biomedical research [2]. The first NHP genome to be sequenced and published was that of *Pan troglodytes* (chimpanzee) in 2005 [3], shortly followed by that of *Macaca mulatta* (rhesus macaque) [4]. These genome assemblies enabled investigation of the origin of human life through comparative genomics and evolutionary analysis to better understand the mechanisms of genetic changes that drive molecular evolution [2–5]. Two key issues in primate comparative genomics concern (i) understanding genomic association using cognitive science to elucidate the mechanisms of human diseases, and (ii) analysis of genomes to uncover the mechanisms of rapid human evolution [1]. Whole-genome sequences for genomes of more than 240 primate species (68% of the total) have now been assembled and are accessible online in the NCBI Assembly database (<https://www.ncbi.nlm.nih.gov/assembly/?term=primate>).

Following successful completion of initial human genome sequencing, we now understand that in addition to the coding portion of the genome (a mere 1–2%), including protein-coding genes that instruct major development and functions, there is also a non-coding portion of the genome with astonishing implications of complexity [6]. The remaining 98–99% non-coding component of the genome comprises highly repetitive sequences and is usually underestimated in the assembled genomes; some parts may contain highly complex repetitive regions that remain undetected despite notable recent developments in sequencing technology. These are considered to be the “dark matter” of the genome [7]. After two decades of improvements, a complete and gapless telomere-to-telomere assembly of a human X chromosome was recently achieved [8]. This milestone has allowed the previous human reference genome (GRCh38) to be updated, and has resolved gaps within the dark matter spanning a novel centromeric satellite 3.1-Mb-long array that currently represents the most accurate and complete vertebrate genome produced to date. The non-coding portion forms the bulk of the genome with highly variable contents from one organism to another, thus making a substantial impact on the variation in genome size (Figure 1a). Such variation may be due to diverse abundances of repeated sequences including tandemly arranged repeats, transposable elements (TEs), and ribosomal genes [9–13]. Comparison of data from the Animal Genome Database [14] (<http://www.genomesize.com/>) shows substantial variation in the distribution of genome size among primate lineages, depending on repeat proportion (Figure 1b). Old World monkeys (family *Cercopithecidae*) and tarsiers (family *Tarsiidae*) have a comparatively larger genome than other lineages, which suggests that expansion of repeat sequences might have occurred during the evolution of their genomes (Figure 1b). Similar cases are observed in other amniote groups, such as mammals and reptiles [13]. Such variation is not correlated with the complexity of the organism. Variation in genome size among species may be associated with different proportions of repeat contents in their genomes [15–17] (Figure 1).

Primate genomes are enriched in repeats (more than 50%), some of which remain uncharacterized [18,19]. Similar to other vertebrates, primate genomes include an abundance of tandem repeats that are organized in such a pattern that the sequences are repeated directly adjacent to each other [20]. These repeat sequences consist of satellite DNA (satDNA), which is defined as tandemly arranged repeats that represent a considerable proportion of the heterochromatic portion of the eukaryotic genome, forming the main structural component (heterochromatin) of chromosomes [13,21–26]. SatDNA has been implicated in a variety of important functions, including segregation during cell division, homologous chromosomal pairing, kinetochore formation, chromatid attachment, chromosomal rearrangements, and differentiation of sex chromosomes [27–33]. Perhaps most importantly, satDNA can constitute rapidly evolving sequences of the genome [34–36] and is now considered to be important in driving genomic and karyotypic evolution [13,22–24,37].



**Figure 1.** Overview of repeat contents in vertebrates and primate genome size variation. (a) Relationship of genome size (C value; red line) with repeat contents (blue bars) in representative vertebrate organisms. (b) Boxplot of the distribution and variability of genome size among primate families. Each dot represents a primate species. Note: Data are sourced from the Animal Genome Size Database (<http://www.genomesize.com/>) [14] and graphics were created in R.

In addition to satDNA, a substantial proportion of primate repeats is categorized as dispersed sequences. These sequences include TEs such as long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs), which are the most abundant interspersed repeats in primate genomes (Figure 2) [38–41]. The evolutionary young (recently emerged) TEs, such as long terminal repeats (LTRs) and the retrotransposons SINE-VNTR-Alu (SVA), may play an important role in the regulation of gene expression and drive evolutionary divergence among primate species [42]. Insertion of new copies of TEs can accelerate the recombination rate; for example, Alu elements may increase the rate of unequal crossover [43].

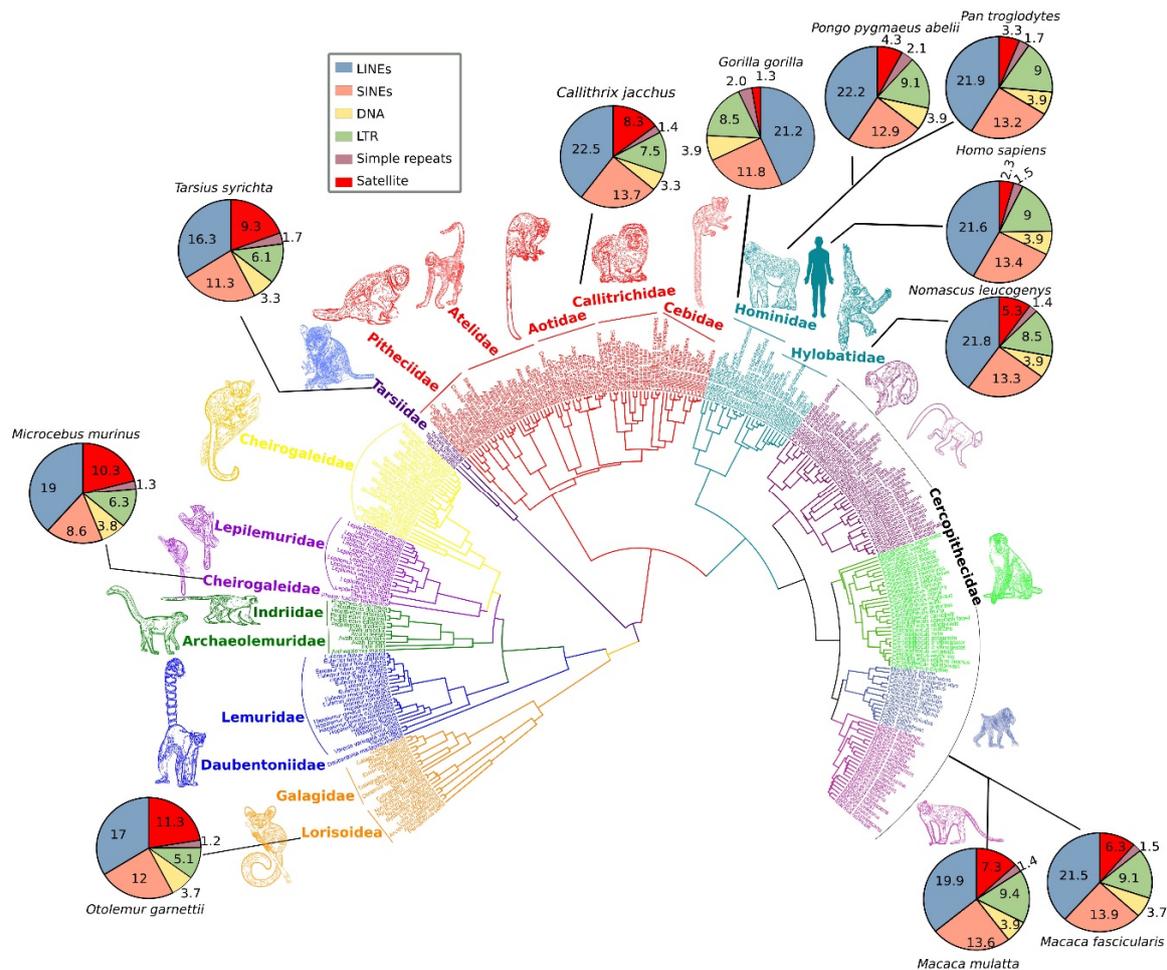
The recent proliferation of new bioinformatics and computational tools has been particularly helpful for the assessment of the variation of repeats. Other developments include advanced next-generation sequencing (NGS) technologies and related software, and the increasing availability of online genomic databases [19,44], all of which allow more reliable testing of models of the functional aspects of the evolution of satDNA within the primate genome. Such analyses offer deep insights into the organization of satDNA within the primate genome and its possible roles in neutral and adaptive evolution. Although primate genomes have been utilized as resources to perform repeatomic analyses, with specific focus on transposable element (TE) contents [45], their satellitomes (complete sets of satDNA families in a genome) remain poorly understood, with just a few recent reports of novel satDNA families [46–49]. Currently available data on chromosomics and molecular and population genetics of satDNA in primates raise many hypotheses concerning their evolutionary origin, expansion in different genomic loci, and functional roles [33,46–48]. To shed further light on these phenomena, we have collected a range of evidence and propose the dynamics of satDNA in primates can be characterized as follows: (i) SatDNA repeats may follow an independent evolution in primate genomes and differences in their genomic abundance among taxa can increase with phylogenetic distance, (ii) the predominant satDNA families are conserved in primates with the exception of certain satDNA types that have undergone extreme divergence, (iii) specific portions of satDNA in the genome show population/species/lineage-level divergence and a paradoxical link with the evolution of centromeres, (iv) the Library model of satDNA evolution is still applicable in primate genome, and (v) satDNA transcriptional activity can mediate regulation of gene expressions that consequently influence wide ranging cellular phenomena. Here, we review and summarize the emerging discoveries of satDNA and discuss its impact on the reshaping of the evolution and dynamics of primate genomes. We highlight different types of primate satDNA and discuss their genomic organization, evolution, and function. We also present an overview of sex chromosomes linked to satDNA and provide detailed insight into lineage-specific divergence of these sequences. We describe an evolutionary model for satDNA and propose a mechanism for the main events that might have occurred during the genomic birth of satDNA. Current challenges in satDNA detection are also briefly discussed.

## 2. Satellite DNA Abundance in Different Primate Lineages

The genomes of most primates, such as monkeys, apes, and humans, comprise up to 50% repeat contents, of which satDNA may constitute as much as 10% of the total number of repeats [50,51]. RepeatMasker data [52] for different primate species indicate that their genomes can contain a highly variable proportion of satDNA (Figure 2). Comparison of these data shows that satellite repeats are highly abundant in certain families, such as nocturnal primates (superfamily *Lorisoidea*), strepsirrhine primates (family *Cheirogaleidae*), and haplorrhine primates (family *Tarsiidae*) (Figure 2), which suggests extensive expansion of satDNA in the genomes of these lineages. By contrast, in *Hominidae* and *Hylobatidae*, satellite repeats are comparatively low in abundance. The genomes of *Hominidae* and *Hylobatidae* are invaded by TEs at higher proportions compared with those of the *Lorisoidea*, *Cheirogaleidae*, and *Tarsiidae* lineages. This observation suggests that phylogenetically close lineages show similar patterns of satellite abundance in their genomes, whereas differences in abundance among taxa increase with phylogenetic distance. The phylogenetic tree of Figure 2 was retrieved from 10kTrees website [53] and designed using Interactive Tree of Life, iTOL [54]. However, data on the relative percentage of satDNA in different primate genomes must be treated carefully, and precise information on satDNA abundance in primate genomes is still lacking due to misassemblies, gaps, and unresolved assembled centromeric regions that span these repeats [19,55].

SatDNAs were initially identified by their buoyant densities (in g/mL) on cesium chloride gradients [56]. This technique was formerly employed for satDNA detection and biased procedures. This technique can identify a single satellite or sometimes multiple satellites in a genome but cannot detect the entire set of satellite families. Modern techniques, such as NGS and fluorescence in situ hybridization (FISH), have replaced traditional methods and have substantially improved detection and

characterization of satDNA [19]. This methodological shift has brought advances in the identification of different repeat types and structural units of satDNA in primate genomes [41,47]. Using cytogenetics, the genomic organization and diversity of satDNA have been widely studied mostly in humans, and to some extent in other primate genomes. As a result, a wealth of knowledge is now available on the localizations of satDNA repeats, their lengths, and different units, variability, and number of copies in different genomes [55,57–63]. These repeats can be categorized as different types of satDNA to better understand their roles, evolution, and applications in phylogenetic analyses. This can include satellites that are generally shared across all eukaryotic lineages and those that are exclusive to primate genomes.



**Figure 2.** A comprehensive phylogeny of 301 primate species based on mitochondrial DNA sequences using Bayesian inference. Pie charts for selected common primate species show percentage differences of repeat types in the respective genomes. The abundance of satellite DNA in primate genomes varies considerably among lineages (red colored area of pie charts). Additionally, the comparative repeatomic landscape shows LINEs and SINEs emerged as the most expanded elements of primate genomes (blue- and orange-colored areas of pie charts) with consistent pattern across diverse lineages. Phylogenetic data were retrieved from the Primates Section of the 10kTrees website (<https://10ktrees.nunn-lab.org/Primates/dataset.html>) [53]. The phylogenetic tree was customized and filtered using iTOL v5 software (Interactive Tree Of Life; <https://itol.embl.de/>) [54]. Different colors represent different clades. Cartoons of the representative primates were drawn using Inkscape software.

### 3. General and Primate-Specific satDNA Types

Certain tandem repeat sequences can be classified by the number of base pairs (bp) into two types as microsatellites (ranging in length from one to six or more bp) and minisatellites (usually

from 10 to 100 bp) [64]. The human genome contains as much as 3% microsatellites [65] and several thousand chromosomal loci enriched with minisatellites [66], also called variable number tandem repeats (VNTRs) [67]. Previous isolation of microsatellites from the human genome has enabled researchers to amplify these sequences in several NHP species, including apes, baboons, macaques, and some platyrrhine monkeys [68–74]. Microsatellites tend to accumulate many substitutions and/or insertions/deletions, and are thus considered to show limited conservation across primate lineages [75]. Many conserved microsatellites, such as AP74, which was discovered in New World monkeys, exhibit similar sequence length (up to 176 bp) in monkeys and humans [76,77]. Boán et al. [78] identified the minisatellite MSH42 in the human genome and performed a comparative analysis in 11 NHP species. Phylogenetic analysis detected several variants of MSH42 and the evolutionary birth of minisatellites in the primate genome was hypothesized. According to this hypothesis, the evolutionary birth of MSH42 took place within an intron early in primate lineage evolution and more than 40 million years ago. Then, various mutations including insertions, duplications, and single nucleotide polymorphism of repeat blocks were probably the major forces governing the generation of this minisatellite and its divergence throughout primate evolution [78]. Certain (TTAGGG)<sub>n</sub> sequences, which are specific monomers of microsatellites, can be repeated multiple times, eventually forming the bulk of the telomeric region up to 15 kb on human chromosomes [79,80]. These telomeric repeats can serve as binding sites for certain nucleoproteins, such as TRF1, TRF2, and POT1, forming a complex termed “shelterin” [81] that interacts with a ribonucleoprotein [82]. This complex is involved in DNA repair processes and the protection against degradation of chromosomal ends [83].

Well-characterized telomeric satellites of the human genome can also be applied broadly as informative markers to study a variety of hominoid species owing to multiallelic variation and a high degree of heterozygosity [70]. The MSH42 locus shows high similarity with immunoglobulin regions and is involved in recombination events as well as in promoting high rates of unequal crossovers [78,84,85]. The telomeres harbor short stretches of sequences termed interstitial telomeric sequences (ITSs), which are located far from the chromosomal ends. To trace the evolutionary origin of these sequences in NHP genomes, 22 ITS loci from the human genome were compared with their orthologs in 12 NHPs, representing species such as great apes, gibbons, Old World monkeys, and New World monkeys. Comparison of sequences indicated that, unlike other microsatellites, these ITS sequences were not derived from expansion of pre-existing TTAGGG monomers but rather emerged abruptly during genome evolution in primates as a result of double-strand break repair [86]. Similar findings were observed from investigation of a chimpanzee-specific ITS. A universal satDNA classification is still the subject of debate; however, most commonly, satDNA can be grouped according to position and association with different chromosomal loci. SatDNA is primarily clustered within the heterochromatin regions of primate chromosomes. The heterochromatic portion is mainly localized in centromeric and telomeric regions, and sometimes within the interstitial regions of the chromosomes [87], whereas satDNA sequences are mostly located in centromeric regions, and the nearby pericentromeres may be enriched with TEs. Different types of primate satDNA are discussed and summarized as Supplementary Table S1.

### 3.1. Centromeric and Pericentromeric satDNA: Primate-Specific Alpha Satellites and HORS

The centromere cores of human chromosomes span abundant and highly enriched stretches of satDNA, and are surrounded by heterochromatin containing a combination of short satDNA sequences and retroelements [29,88]. Occasionally, these centromeric regions are termed “satellite centromeres” [89]. The centromere is an important region of the chromosome for preservation of genetic materials and plays a critical role in chromosome segregation, cell division, kinetochore organization, and spindle attachment [89–92]. In primates, the bulk of the centromere is composed of the pancentromeric alpha satellite (AS), organized as stretches of 171 bp monomers in a head-to-tail fashion extending for ~250 kbp up to ~5 Mbp per chromosome [93–96] (Figure 3a(i)). This structure has been reported across diverse groups, including great apes, Old World monkeys, and New World

monkeys [96–102]. These centromere-associated satellites are arranged as superfamilies (SFs) that can be orthologous between human and gorilla [60]. The surrounding pericentromeric satDNA are essential elements that assist in stabilization of DNA–protein binding and regulation of chromosome segregation [58,61]. These pericentromeric satellites vary greatly across NHP species but can be conserved among closely related species or may be species-specific [20,103]. For instance, a large block of human chromosome 9 that spans a pericentromeric area enriched with satellite III (SatIII) shares close homology with the gorilla sequence [104]. The Y chromosome of NHPs may carry higher numbers of copies of satellite III sequences than the human Y chromosome [105]. FISH mapping of the pericentromeric-type satellite pW-1 SatIII DNA on chromosomes of various NHP species showed that these sequences might be lacking in the genomes of squirrel monkey (*Saimiri sciureus*) and baboon (*Papio hamadryas*) [105]. These centromeric satellites can vary substantially across different species, but certain species-specific or even highly conserved satDNA may also be present in the centromere domains [20,103]. For example, two major families of centromeric satellites, termed C1 and C2, detected in Old World monkey species crested mona monkey (*Cercopithecus pogonias*) and sun-tailed monkey (*Cercopithecus solatus*) have remained highly conserved [48]. For Old World monkeys, apes, and humans, each genome harbors evolutionarily distinct AS monomers [106]. Although most primate centromeres can be enriched with satellite repeats, there are certain chromosomes of orangutan that comprise non-repeated centromeres [92,107–110]. In such cases, the centromeres may resemble newly formed neocentromeres as a result of disruption in the centromeric region, such as in humans [92,111]. Such non-repeated centromeres are likely to be evolutionary new centromeres (ENCs), forming neocentromeres that might have subsequently gained repeat sequences to stabilize the genome and become fixed in populations. This phenomenon can also occur in the centromeres of several non-primate species, such as horse and chicken [112,113]. In the following, we focus mainly on the predominant centromeric satDNA in primate genomes as AS repeats.

The AS repeats were first observed as tandem repeats in the African green monkey (*Chlorocebus aethiops*) genome [93], followed by identification of homologous repeats in New World monkeys and apes [96,114]. These sequences are considered to be critical components for the various functions of primate centromeres [94]. Previous results suggest that AS sequences were involved in stabilization of ENCs after their emergence in primate genomes [109,115]. Human and macaque chromosomes contain a total of 14 ENCs, of which nine ENCs in the macaque genome show abundant arrays of AS [109]. Interestingly, ENCs occur in macaque chromosome 4 and human chromosome 6, which are orthologous to each other (Figure 3a(ii)) [109,116,117].

The AS monomer size is 171 bp, tandemly arranged in a head-to-tail manner, and shows as much as 70% sequence similarity. The combined monomers can form a long array spanning an uninterrupted 250–5000 kb stretch of repeated satellites, giving rise to high-order repeats (HORs) (Figure 3a(iii)). A certain monomer in the HORs with a sequence size of 17 bp is termed the CENP-B box. This motif acts as a protein-binding site for a centromeric CENP-B protein in primates. The human genome project, which was declared complete in 2003, was still unable to recover a large proportion of the centromeric and other repeats, including more than 10% of the contents of the whole genome, mainly sex chromosomes. However, subsequent technological developments enabled assembly of the entire human Y chromosomal centromere [62,118]. The Y chromosome assembly could be used as a reference sequence to extend evolutionary insights into the centromeric repeats of NHPs for which Y chromosome assemblies have not been hitherto accomplished.

In primates, the flanked regions of centromeres have specialized HORs arrays, whereas AS sequences are organized as non-structured and heterogeneous repeats, forming distinctive pericentromeres. In these pericentromeres, AS sequence repeats are arranged as monomers instead of HORs and are interrupted with additional elements, mainly retrotransposable elements in humans [119] (Figure 3a(iii)), which may also be common to other primate genomes. The pericentromeres of certain human chromosomes may also show enrichment of several other repeat sequences, including the 5 bp satDNA II and III type sequences [103,120]. The AS sequences can show nucleotide variation when one

monomer is compared with the repeats of the same array, with nucleotide identity ranging from 70% to 90%. The sequences of a monomer in one array may show up to 95% similarity with its counterpart unit in the other array at the same locus [63,121,122]. In the human genome, the organization of HORs with their monomer units has been extensively studied [65,97,123,124], and shows the occurrence of various subfamilies of chromosome-specific AS sequences. The sequences of HORs in great apes, such as orangutan, gorilla, and chimpanzee, show a lower degree of variation in comparison with HORs observed in the human genome [125–128]. Initially, it was presumed that the organization of HORs might be restricted to hominids; however, HORs were subsequently detected in the genomes of gibbons [101,102,129] and of Old World and New World monkeys [48,102,130]. During the evolution of the primate genome, the 170 bp AS monomer underwent a series of sequence variations [87]. A novel AS monomer type of 189 bp was discovered in the centromeres of gorilla [60]. Chromosome-specific subfamilies are absent in Old World and New World monkeys as well as in gibbons [87,101,106]. Cloning, sequencing, and hybridization of acrocentric chromosomes revealed novel AS sequence repeats in Azara's owl monkey (*Aotus azarae*), which is a species of New World monkey [22,23]. These repeats include three megasatellites, namely OwlRep, OwlAlp1, and OwlAlp2, which vary in size from 184 to 344 bp as identified in the centromeric and pericentromeric regions. Analysis of retina samples using three-dimensional FISH revealed that OwlRep is the major component of heterochromatin, which indicates its role in the evolution of night vision in this species [131,132]. Recently, Cacheux et al. [49] investigated the evolutionary dynamics of AS sequence repeats and their diversity in the Old World monkeys *Cercopithecus pogonias* and *C. solatus* using targeted sequencing and FISH mapping. These authors reported evidence of chromosome-specific subfamilies that might have evolved through homogenization. The OwlRep repeat shows ~82% homology with a satellite sequence termed HSAT6, which is a 126 bp long tandem centromeric repeat. The HSAT6 sequence was also detected in the owl monkey genome, and comparative analysis revealed its broad distribution among hominoids and New World and Old World monkeys. Phylogenetic analysis confirmed that OwlRep evolved from HSAT6 [132].

In addition to AS, an additional type of satellite family termed the beta satellite is distributed in the heterochromatin of primates [133–135]. Beta satDNA are repeats that comprise ~68 bp monomers. They are predominantly organized in the shorter arm of acrocentric chromosomes and arranged in stretches several kb in length [136–139]. The beta satDNA repeats can form complexes with arrays of specific repeats, termed D4Z4 repeats, at certain acrocentric loci, such as 10q26 and 4q35 [140,141]. Evolutionary analyses involving cloning and FISH experiments have predicted that 4q35 containing D4Z4 repeats might represent an ancestral locus with an extensively radiated sequence region that evolved after the divergence of hominoids and Old World monkeys [142–144]. The origin and evolution of beta satDNA vary in diverse species of hominids, such as humans, chimpanzee, and gorilla [145,146]. FISH mapping data confirm that D4Z4 is also conserved in Old World and New World monkeys, whereas in primates distantly related to humans (e.g., lemurs), this sequence has retained tandem repetition but conservation is limited to promotor regions [147]. Genomic analysis of orangutan has revealed the origin of beta satDNA in earlier ancestors of hominoids and shows that these repeats are preferentially located in pericentromeres [135]. This study concluded that these repeats originated as low copies, remained non-duplicated in the early ape ancestors, and later evolved as duplicons acquiring the typical characteristics of classical satellites in humans and other primates. Adjacent to ASs, the classical non-alphoid satDNA repeat families I, II, and III are located in pericentromeres of human chromosomes [95]. The human genome includes the Sat III family, which is composed of GGAAT and GGAGT repeat sequences in different percentages. The satellite III family is mainly localized on the short arm of acrocentric chromosomes in humans and other primate species. This family is also present in the chimpanzee, gorilla, and orangutan genomes [148,149]. The chromosomal organization of this satellite family has provided interesting evolutionary insights into primate genomes [105]. Sequence comparisons have detected variation across different primate species and suggest that the Sat III family might have appeared ~16–23 million years ago in Hominoidea [105]. The evolutionary origin and

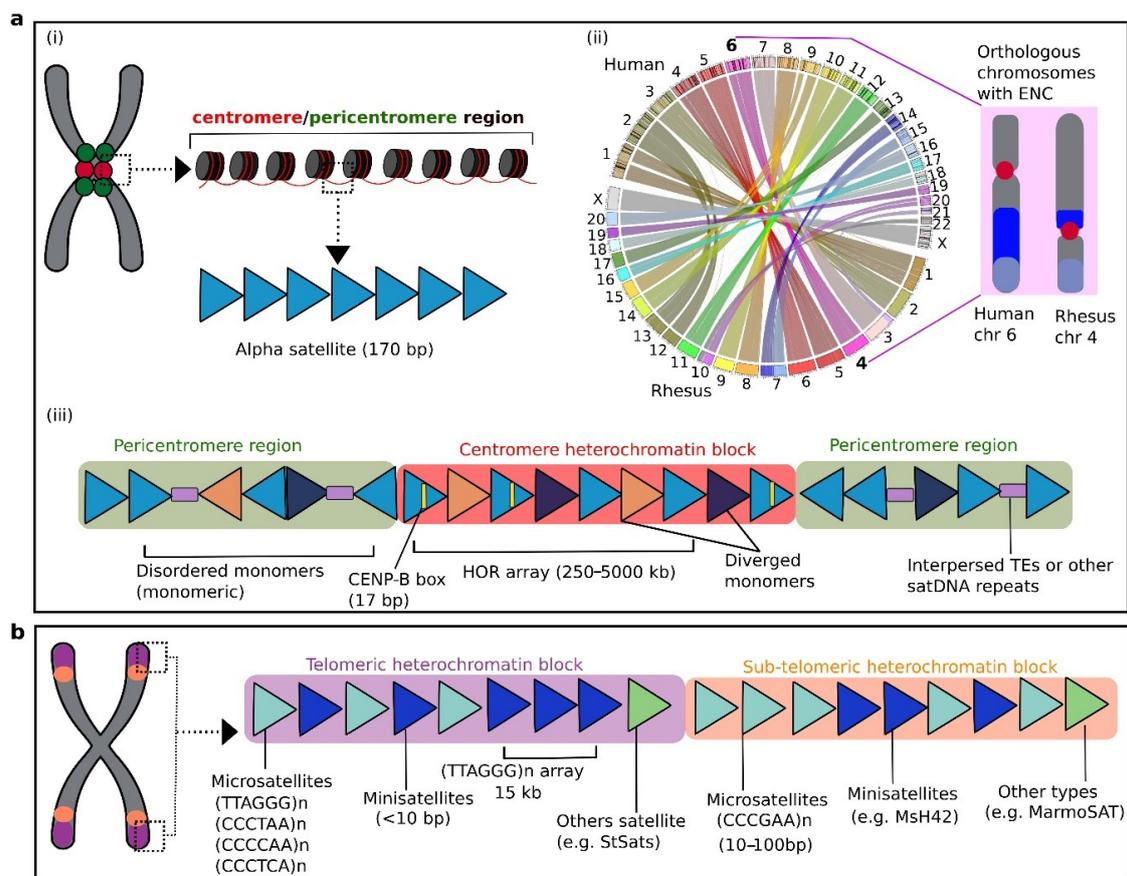
extensive diversification of centromeric satellites in primate genomes remain unclear; however, it is speculated that TEs are the possible progenitors and sources that form novel satellites by insertions into existing satellite regions [119].

### 3.2. Telomeric and Subtelomeric satDNA

The telomere is located at the end of the chromosome and is enriched with a non-coding, repetitive DNA sequence. The 500 kb region of each chromosomal arm terminal is the so-called subtelomeric region [150]. Both telomere and subtelomere have high-density of satDNA repeats. Telomeric regions of the primate genome show a high frequency of minisatellites, which also occur in other loci of chromosomes [67,151]. The bulk of telomeric-specific regions are mainly composed of (TTAGGG)<sub>n</sub> microsatellites in humans [79]. Adjacent to the telomere, the subtelomere region is mostly enriched in rapidly evolving satellite repeats with variable levels of repetitiveness and size [57,152,153]. Although these subtelomeric satellites can be species-specific and often chromosome-specific, there are also satellites that remain highly conserved [154]. The microsatellites (CCCTAA)<sub>n</sub>, (CCCCAA)<sub>n</sub>, and (CCCTCA)<sub>n</sub> are present in telomeres of primates [155], whereas (CCCGAA)<sub>n</sub> is restricted to subtelomeres [156] (Figure 3b). In New World monkeys, the subtelomeres can carry novel satDNA sequences. The subtelomeric regions of callitrichid monkeys harbor a satellite termed MarmoSAT that is composed of a 171 bp motif [157]. The MarmoSAT occurs as a monomer, whereas in common marmoset (*Callithrix jacchus*) it is organized in HORs with a sequence of 338 bp. Recently, some intriguing groups of satDNA sequences enriched with AT nucleotides, termed StSats, have been reported in telomeres of humans and great apes, including bonobo, chimpanzee, gorilla, and orangutan [47]. The StSats are located in proximity to telomeric regions [158–160]. Astonishingly, these satellites are very highly enriched in the gorilla and chimpanzee genomes compared with their abundance in humans [47]. Previously, it was hypothesized that these repeats occurred in hominid ancestors and were lost in humans [158–160]. The abundance of StSats repeats in the bonobo, chimpanzee, and gorilla genomes indicates that these sequences might contribute to important genomic functions in these species. Different functions have been proposed for these repeats that include their role in meiosis, telomere clustering, and control of replication duration with telomeric regions [158–160].

The distribution of two distinct satellite repeats, termed Cap-A and Cap-B, was reported in a New World monkey species, *Cebus paella* [161]. The Cap-A sequence is 1500 bp long and forms heterochromatic blocks in the interstitial sites of chromosome 11 and a few telomeric regions. This suggests that this sequence underwent a new episode of amplification in New World monkeys. This satDNA repeat is absent in most marmoset species and present in species of the family *Cebidae*. By contrast, the Cap-B satellite, which is 342 bp long, is mainly localized in the centromeric regions of many chromosomes of New World monkeys. The Cap-B monomer sequence shares more than 60% identity with AS repeats, which indicates that Cap-B might be the New World monkey homolog of Old World monkey AS repeat sequences. Telomeric satDNA sequences can participate in the formation and maintenance of telomeres, and may have an incidental role in cases losing of conventional telomeric repeats. In this way, telomeric ends are stabilized by satDNA [162]. Further, it has been demonstrated that telomere-like sequences interspersed within subtelomeric DNA may also play a role in subtelomeric recombination and transcription, via alternative lengthening of the telomere pathway and in telomere healing [163]. It is necessary to identify and characterize the telomeric/centromeric satDNA sequences particularly at the breakpoint sites because of their role in mediating chromosomal rearrangements [164] that occurred during primate evolution. Such analyses have been performed with regard to the gorilla-specific translocation [165] as well as the chromosome scale variations that serve to distinguish human and chimpanzee chromosomes. Various hotspot rearrangement regions of the gibbon genome have also been characterized [166]. In contrast to the great apes, gibbons have chromosomes with higher levels of rearrangement compared to ancestral primate karyotypes [167]. A comparison of human and chimpanzee karyotypes showed that two ancestral chromosomal homologs of chimpanzee chromosomes 12 and 13 underwent a fusion event to give rise

to human chromosome 2 [168]. This fusion was mediated by recombination between telomeric satDNA repeats of the two sub-metacentric ancestral chromosomes. The hyper-expanded repeats are localized in subtelomeric regions of chimpanzee chromosomes. These repeat enriched regions are also prone to other types of rearrangement events such as duplicative transpositions and inter-chromosomal sequence variations [169]. Since many primate-specific rearranged loci are enriched with high-copy repetitive sequence elements such as alpha satDNA repeats, SINEs, LINEs, and LTRs, a range of different molecular mechanisms were probably involved in promoting chromosomal breakage during the evolution of primate genomes [164]. Genome-wide scale analyses at higher resolution are necessary to determine the precise mechanisms underlying the different types of rearrangement, and to assess their relative contribution to the process of evolutionary change.



**Figure 3.** Schematic illustration of satellite DNA repeats and their organization in primate genomes. (a) (i) Primate centromeric (red) and pericentromeric (green) regions are enriched with alpha satellite (AS) DNA as the most abundant satellite repeats of primate genomes and form the bulk of the heterochromatin core. (ii) A sketch highlighting the orthologous chromosomes and centromeric repositioning as evolutionary new centromeres (ENCs) between human and rhesus macaque. The circos plot depicts the syntenic relationship between the two genomes. Circos graphics was plotted using Synteny Portal [117]. Note that human chromosome 6 is completely orthologous to macaque chromosome 4, with evolved centromeres [109,116]. (iii) The AS constitute the tandem repeat units (blue triangles) and can be either organized as disordered arrays (monomeric) mostly located in pericentromeres, or highly ordered in a head-to-tail fashion (HORs) forming longer arrays in centromeres. Some monomers may also have a short sequence termed the CENP-B box (yellow line), which binds the centromeric regions to the DNA-binding proteins. Diverged monomers (orange and dark triangles), and interspersed repeats (purple rectangles) are also depicted. (b) Telomeric and subtelomeric regions of primate chromosomes are enriched with distinct microsatellites (light blue) and minisatellites (dark blue). Various primate-associated satellite examples are shown.

#### 4. Sex Chromosomes: A High-Impact Arena for satDNA

In addition to the distribution of satDNA repeats on autosomes, these sequences are specifically enriched in distinct loci of particular chromosomes such as microchromosomes, supernumerary chromosomes, and sex chromosomes [13,170,171]. In primates, highly heteromorphic sex chromosomes have evolved from a pair of autosomes [172]. Divergence and erosion of the Y chromosome resulted in loss of several functional genes and accumulation of different repeats, including TEs and satellites [173–177]. FISH mapping of the SatIII family in different primate species revealed that these satellite repeats occur on the Y chromosome of humans, chimpanzee, gorilla, orangutan, and gibbons [105]. Weak hybridization signals on the human Y chromosome indicate fewer copies of these repeats than in other primate species. FISH mapping of a subtelomeric satellite, MarmoSAT, in four species of marmosets confirmed the occurrence of these sequences on both X and Y chromosomes [157]. MarmoSAT is localized to the short arm, whereas the telomeric repeats (TTAGGG)<sub>4</sub> are distributed on the long arms of the sex chromosomes of *Callithrix penicillata* and *C. geoffroyi*. RNA-sequencing analysis further revealed that MarmoSAT is transcriptionally active with higher levels of expression in spleen, thymus, and heart tissues, which suggests it may play a role in telomeric chromatin [157]. An additional sex-linked satDNA is gamma satellite, which comprises 220 bp repeated GC-enriched units and is mostly embedded with AS sequences to form repetitive clusters [178–180]. Gamma satellite is mainly localized in pericentromeres of the X and Y chromosomes in humans [28,179]. An interesting satellite of the human Y chromosome is HSAT 1, which occurs in the heterochromatin [181]. This HSAT 1 is surrounded by palindromic AT-rich repeats and an AluSc element. Together, these three units form a sequence with a tripartite structure. These repeats are also localized on the autosomes but not on the sex chromosomes of NHPs. Exclusive occurrence on the human Y chromosome shows that this sequence has been duplicated on the Y chromosome after divergence of the human genome from its common ancestor [181]. Characterization of the diversity of repeat sequences has resulted in the detection of AS sequence subfamilies on the Y chromosome of the Old World monkey *C. solatus* [48]. The C3 and C4 satellites are localized on the Y chromosome with high copy numbers but are almost absent in autosomes. The Y-specific organization of these satellite dimers (C3 and C4) emphasizes that the Y chromosome does not undergo recombination. The Y-linked satellite repeats in many primate species have remained unexplored, except in gorilla, chimpanzee, and humans, for which Y chromosomes have been almost completely assembled [176,182,183].

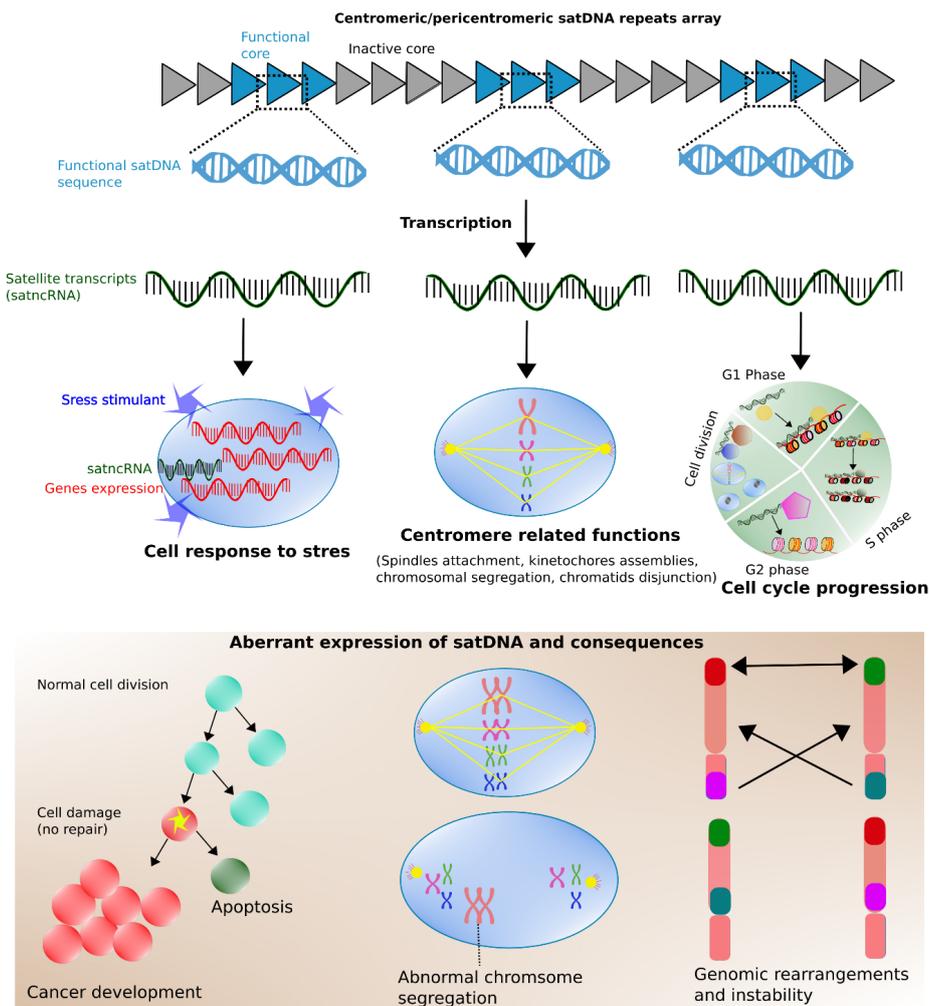
An advanced methodology was proposed to identify Y-linked sequences in humans [184]. This technique located 119 scaffolds of ~18 Mb present on the human Y chromosome. A total of 34 sequences of the 119 scaffolds constitute a higher percentage (74%) of repeats, mainly including Y-linked satellites. These satellites map to the centromere and Yq12 band of the Y chromosome. Recently, short- and long-read sequencing analysis has uncovered male-biased satellites among great ape species [47]. Differences in density between male and female repeats resulted in identification of 18 satDNA sequences that included the human Y-specific “(AATGG)<sub>n</sub>” sequence with its three types of satellite units: DYZ1, DYZ17, and DYZ18 [165]. These three satellites have also been mapped on the Y chromosome of bonobo, chimpanzee, gorilla, and orangutan [105]. Different StSats sequences have been identified on the Y chromosome of bonobo and gorilla [47]. These studies have greatly improved current knowledge of satellite contents of the Y chromosome in apes. Differences in abundance or enrichment of unique satellite repeats were previously considered to be a primary step in distinguishing the X and Y sex chromosomes [185]. An emerging hypothesis states that the composition of Y heterochromatin may differ from the remaining chromosomes owing to several factors that include lack of Y recombination, the putative role of heterochromatin in silencing of Y, and the smaller effective population size of Y [175,186,187]. Apart from the Y chromosome, the human X chromosome is among the best studied sex chromosomes among primates. FISH mapping revealed the abundance of a type of HORs termed DXZ1, which is located in the functional centromere of the X chromosome [29]. The accumulation of HORs in the X and Y chromosomes has been linked to topoisomerase II cleavage activity detected in active centromeric regions [188–190]. Functional

and genomic analysis has further unveiled many interesting findings on the pericentromere of the human X chromosome, which is why the X chromosome has emerged as an exciting model system for investigation of the centromere [29]. The particular occurrence of DXZ1 HORs proximal to the shorter arm of the X chromosome is especially important to gain advanced knowledge of the function of the centromere. A survey of satellite repeats in several primate species compared sequences in the pericentromeric region of the X chromosome and offered new insights on the evolution of the centromere [30]. The evolutionary analyses showed the addition of new sequences into centromeric regions as a series of punctuated events including frequent homogenization and inter-chromosomal exchanges of monomeric AS monomers in early primates. Phylogenetic analysis of X-linked AS monomers detected certain domains containing recently evolved LINE retrotransposons flanked with these satellites. In-depth evolutionary comparison of this junction revealed a striking conservation of AS sequences, which supports its ancestral nature in primates such as baboon, chimpanzee, gorilla, orangutan, macaque, and vervet monkey. In addition, this comparative analysis demonstrated that the centromere of the X chromosome in primates might have evolved as a result of expansion events of repeats. Among the most studied X-linked satellites is the macrosatellite, DXZ4, which consists of 3 kb repetitive units at the Xq23 position [191]. FISH mapping has revealed the organization of DXZ4 on the X chromosomes of chimpanzee, gorilla, and orangutan [192]. These sequences were subsequently mapped on the X chromosomes of male and female rhesus macaque *Macaca mulatta* [48]. Southern blot analysis revealed hybridized fragments of different sizes (50–350 kb), which points to its possible VNTR nature in primates, including great apes and Old World and New World monkeys [147]. Several important features, including the promotor, CpG sites, GC content, and CTCF binding site associated with the DXZ4 satellite have been explored in phylogenetically distant primate species. Integrative genomic approaches, coupled with chromatin conformation (immunoprecipitation and immunofluorescence) experiments in macaques, further indicate that DXZ4 is organized in heterochromatin on the active X chromosome, whereas this sequence is packaged in euchromatin on the inactive X chromosome [147].

## 5. Transcription of Satellite Repeats: Hidden Switches for Dialing Gene Expression Up and Down

Satellite repeats have long been regarded as junk DNA [193]. However, there is increasing evidence to suggest the functional importance of these sequences. The transcriptional activity of satellites is a well-known feature in diverse species [194]. SatDNA sequences have been determined to be involved in various functions, such as developmental processes, stress response, cell proliferation, and cancer [195] (Figure 4). In principle, satellite transcripts are most likely associated with important genome functions, for example, centromere structure, kinetochore assemblies, and chromosomal segregation [58,87,195]. Recently, applications involving new methods of sequencing and bioinformatic analysis have enabled investigation of the functioning of satDNA in primates [196]. Alpha satellite non-coding RNA (satncRNA) has been detected in a prenucleosomal complex that contains CENP-A and HJURP [197]. During the G1 phase, RNA polymerase II associates with chromatin fibers when CENP-A and HJURP bind with a 1.3 kb satncRNA, which assists in localization of CENP-A on centromeric chromatin. The absence of satncRNA has been linked with several mitotic defects and decreased levels of CENP-A recruitment, thus emphasizing its important role in centromere-related functions. In addition, in human cells satncRNA has been associated with heterochromatin formation by recruiting an enzyme termed SUV39H that can bind with the pericentromere [198]. Other crucial functions of AS repeats include regulation of spindle attachments, maintenance of heterochromatin, and disjunction of chromatids (Figure 4) [198–202]. These phenomena might be regulated by the association of satncRNA transcripts with AURORA B and SGO1 proteins [199], which might further associate with SUV39H1 proteins and promote heterochromatin stability [198,200,202]. It was previously believed that the heterochromatic portion of chromosomal regions are transcriptionally silent; however, satellite sequences in centromeres are actively transcribed during cell division [203,204], which plays an important role in kinetochore preservation and centromere cohesion [205,206]. The satDNA transcripts can perform different key

functions that appear to be linked with their chromosomal locus. Transcripts of pericentromeric satellites are involved in the formation of chromatin. The SatIII transcripts can play an important role during the cellular response to stress in humans. In particular, heat shock factors [207,208] can initiate SatIII transcription and form nuclear stress bodies surrounding SatIII loci [209]. This can further cause splicing of stress response-associated genes, subsequently downregulating their transcription and preventing stress-induced apoptosis and subsequent cell death [210]. The SatIII transcripts are not only associated specifically with thermal stress response, and have been detected in the absence of stress [211]. Other functions of satDNA transcripts are kinetochore assembly, regulation of telomere capping and elongation, epigenetic control of heterochromatic regions, and gene expression [58,205,212].



**Figure 4.** Diagrammatic summary of satellite DNA transcription highlighting various functional roles in cellular processes. The most studied transcripts of satellite repeats are those localized in the centromeres. The centromere/pericentromere core contains the active regions with satellite sequences that can be transcribed into satncRNA. These satncRNAs are associated with various functions. For example, during cellular stress, the satncRNA can regulate the expression of important genes, such as HSF1 (Heat Shock Factor 1), to produce nuclear stress bodies. In addition, satncRNA can also regulate the splicing of associated genes that are vital in stress responses. More importantly, satncRNA transcripts have been linked with centromere-related functions and cell-cycle progression. During the G1 phase, the satncRNA can facilitate the loading of CENP-A (yellow circle) at centromeres, which is distributed to every daughter strand in the S phase. In the G2 phase, the satncRNA transcripts form associations with SUV39H1 (purple pentagon) before initiation of cell division. During mitosis, satncRNA binds with SGO1 and AURORA B proteins, and assists in kinetochore assembly, spindle attachment, and chromosome segregation-related functions.

The genomes of great apes contain three different types of AS monomers such as (AATGG)<sub>n</sub>, (TTAGGG)<sub>n</sub>, and AT-rich 32-monomer satellites. These sequences are present in abundance and are particularly interesting because of their function in great apes. The (AATGG)<sub>n</sub> satellite, which is the source of the human HSAT2 and HSAT3 satellites, is transcribed into a long non-coding RNA that is important in thermal stress response [210]. This satellite also occurs in orangutan but its variability remains unknown in many ape species. The satellite (TTAGGG)<sub>n</sub> localized in the telomere is critical in aging, cell division, and genome stability [213,214]. Another subterminal satellite (StSat) repeat, consisting of 32-bp-long AT-rich units, occurs in the proximity of subtelomeric regions in bonobo, chimpanzee, and gorilla, and is involved in telomere metabolism [158,160,215–218]. An irregular StSat expression level has been increasingly linked with cancer development [195]. The overexpression of satellites during stress can be analogous to cancer because certain features, such as abnormal chromosomal segregation, aneuploidy, or reduction of chromatid cohesion, prevail in both states. Hypomethylated pericentromeric DNA and up-regulation of satellite transcripts have been reported in cancer [219,220]. The highly repetitive nature, duplicated copies, and multiple genomic positions make transcriptional analysis of satDNA sequences a challenging task. The commonly available short-read NGS technologies show substantial limitations for investigation of satellite transcripts and their expression dynamics. A major shortcoming of short-read sequences is their inefficiency for assembly of the transcripts of large repeats. The recent advent of ultra-long-read sequencing may overcome this limitation in the near future. Transcriptional analysis of satellite repeats, under normal cellular functioning and in response to disease infection, has become an intriguing focus of ongoing research. However, much remains unclear about the expression of these sequences, and methodological improvements are needed to attain an improved understanding of their transcript functions.

## 6. Species and Population-Specific Variation: An Auspicious satDNA Feature for Genome Evolution

The evolution of satellite repeats in primates is extremely complex because some have remained conserved throughout evolution while others exhibit dynamic variation within the same population [221]. Heterochromatic sequences show remarkable interspecific variability in structure and size. This variation has also been detected between phylogenetically close primate species. The long arm of the human and gorilla Y chromosome carries heterochromatin as a major component, whereas heterochromatin is almost absent in chimpanzee [222]. Intraspecific variation in heterochromatin across human populations have been investigated intensively [62,223]. A Y-linked satellite of size 3.6 kb, namely HSAT3 in humans, contains a DYZ1 sequence that shows substantial variation among individuals and populations [223]. Similarly, centromere repeats of human X chromosomes are highly variable in sequence length among populations [62]. Certain human populations have shorter (15 kb) arrays within the heterochromatin of the neocentromeres that can result in defective sister-chromatid cohesion [224].

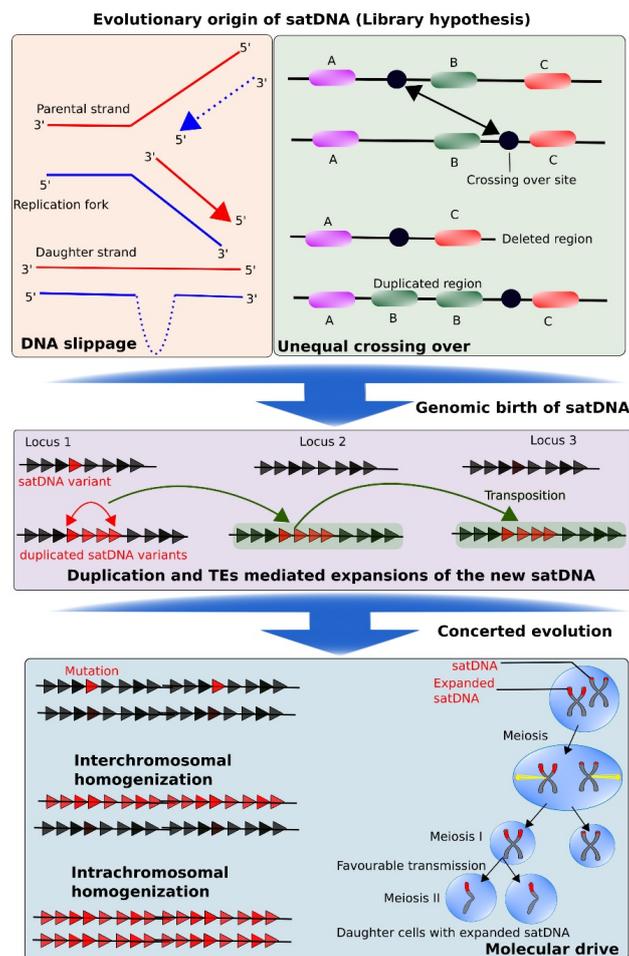
Sequence composition analysis of the SatIII DNA subfamily Pr-1 revealed up to 4.5% sequence variability in gorilla compared with the human sequence [105]. Members of the Pr-1 subfamily are absent from the chimpanzee genome either because the sequence has been lost after the divergence of gorilla and chimpanzee from the common ancestor or because the subfamily emerged independently in the gorilla and human genomes. In great apes, nucleotide variation of satellite repeats was investigated using short- and long-read sequencing data, and repeat densities were measured and compared among ape species. Interestingly, satellite density varied considerably among ape species. At a certain level of abundance of the satellites in the genome, the satellite variants can become dominant, and subsequently result in intrachromosomal homogenization in species and formation of chromosome-specific arrays (Figure 5). Recently, chromosome-specific variation of AS repeats in human populations has been revealed [62,124,225]. Three AS sequences, namely D17Z1, D17Z1-B, and D17Z1-C which are localized adjacent to the centromere of human chromosome 17, show high variations [124]. These satDNA repeats have been widely investigated in a number of recent studies [226–229]. Each of the three aforementioned satellites can serve as a functional centromere to recruit CENP-A histone

proteins; therefore, the functional multiple AS sequences located on a single chromosome are termed epialleles [230]. Approximately 70% of the analyzed population harbors assembled centromeres at the 16mer D17Z1 position, whereas the remaining 30% displays differential assemblies of centromeres at the D17Z1 and 14mer D17Z1-B loci. Centromere assembly is supported by the D17Z1-B epiallele in human artificial chromosomes, but to date no human individual with this allele has been detected. Therefore, it is speculated that individuals homozygous for the D17Z1-B epiallele may represent rare but viable variants in the human population. Likewise, human sex chromosomes have been explored by NGS analysis and several satellite sequences with X- and Y-linked variants that differ in size have been detected [62]. Ongoing research by several groups has built upon this foundation to understand satellite repeat variation [231,232], and further indicates that HOR variation in the centromeres might not be specific to human chromosome 17.

Over the course of the past two decades, AS sequence divergence has been the main focus of studies of satellites in primates. Array patterns of AS sequences in African green monkeys were identified with 1–5% divergence among inter- and intrachromosomal monomers [233,234]. Other primate genomes possess repeated units with 30–50% divergence in monomers among species, although the complete array exhibits only 1–10% variability [98,235–239]. Alexandrov et al. [100] reviewed and characterized the divergence and ancestry of AS in lower primates. The ancestral AS monomer, termed S1, may have acted as a source sequence to give rise to a S1–S2 dimer repeat, which is a typical sequence of Old World monkey genomes. This S1–S2 dimer is formed by combination of S1 with its S2 variant, followed by duplication during the evolution of lower primates. Conversely, the S3–S4 dimer, which is characteristic of New World monkeys, evolved by amplification of S1-like variants with different divergence rates. In certain species of the New World monkey genera *Pithecia* and *Chiropotes*, a trimer unit (combination of the S3, S4, and S5 monomers) is formed as a result of unequal crossing over between S3–S4 dimer repeats. The S5 monomer in this trimer unit evolved from S3 and S4, and shares their sequences. Each of the aforementioned HORs is located on several chromosomes forming the bulk of heterochromatin in the corresponding primate species. In some Old World monkey species, such as *Macaca fuscata*, the certain repeat units are almost identical with those of *Chiropotes* and *Pithecia*, including S1–S2 dimer repeats that remain conserved throughout the genera *Papio* and *Macaca* [100]. Some primates, such as great apes, have acquired novel monomers that have been duplicated and intermixed with older monomers [240–242]. As a result, three new suprachromosomal families (SF1, SF2, and SF3) emerged in great ape genomes. In humans, the new AS families are clustered in the centromeres of all autosomes but are absent on the Y chromosome, whereas in other apes they are distributed on almost every chromosome. Each satellite family represents a distinct chromosome-specific structure that is defined by 2–30 monomeric HORs arranged tandemly with more than 95% identity [240]. Each family can comprise thousands of copies of HOR arrays that can span 250–5000 kb sequences (reviewed in [125]). Certain AS families in humans that evolved into the new SFs can contain both ancestral and new monomers. These ancestral sequences might have amplified and accumulated mutations during their insertion or relocation into nonhomologous centromeres and other chromosomes. Owing to subsequent amplification of mutated sequences, a novel HOR could have been formed that now contains divergent copies of ancestor repeats eventually forming a new array [97,242].

Mutations (polymorphism) can occur in primate satellite sequences at both single bases, i.e., single-nucleotide polymorphism (SNP), or multiple nucleotides (structural variants), which may involve segments of HORs [243]. As described earlier, several cases have been reported in primates that show sequence variation in HORs to be species-specific, population-specific, or even chromosome-specific with different divergence rates. Most of these variations have been studied in AS repeats of humans [50], although divergence repeats have also recently been investigated in several NHPs using modern sequencing techniques [48,244]. However, these studies tend to focus on interspecific comparison, and population variability of satDNA sequences remains incompletely understood in NHPs. Research on satDNA sequence divergence or variation may provide further interesting genomic insights into the evolutionary processes among lineages or populations. Analysis of genome sequences has resulted

in the characterization of a 187 bp megasatellite sequence named OwlRep, which is distributed in the heterochromatin of simian primates (infraorder Simiiformes) and owl monkeys [245]. Interestingly, primate infraorder Tarsiiformes and suborder Strepsirrhini have only one copy of HSAT6, whereas several species of infraorder Simiiformes can carry many copies. Comparative sequence analysis of these copies revealed duplication of HSAT6 in New World monkeys and demonstrated that OwlRep probably originated during the divergence of owl monkey lineages from New World monkeys. In addition, species-specific variation in AS monomer size is reported among New World monkeys (sequence length from 340 to 350 bp) and hominids (171 bp). The CENP-B box was identified in the genome of three New World monkey species, including squirrel monkey (*Saimiri sciureus*), tamarin (*Saguinus oedipus*), and marmoset (*Callithrix jacchus*) [244]. The CENP-B sequence of each species not only varied in length but was also located at different chromosomal positions.



**Figure 5.** Model of satellite DNA evolution. Genomic birth of satellite DNA can occur as a result of different mechanisms. According to the Library hypothesis, the two main proposed phenomena are DNA replication slippage and unequal crossing over, which can cause mutations and de novo formation of satellite DNA (variants shown as red triangles). The newly formed satellite region can undergo several duplications and subsequent transposition events that expand the new satellite throughout the genome. Transposable elements mediate the spreading of newly evolved satellite repeats to different loci. This is followed by cohesive evolution of the genomic region to homogenize the entire array through selection. Finally, the evolved satellite repeats are established by sexual reproduction. The chromosome region with expanded satellite is inherited preferentially via a molecular mechanism known as “drive”. Two homologs have the same satellite DNA (red) but with a larger centromere, and one homolog with expanded satellites is attracted by spindle fibers and driven to the daughter cells.

## 7. Evolutionary Birth and Expansion of Satellite DNA

As described earlier, satDNA repeats are highly variable sequences that represent species- or genus-specific genomic fractions and reflect trajectories of short-term evolutionary changes [21,246–248]. Nevertheless, the significance of these sequences in studying genomic functions and structuring during different evolutionary events, together with increasing data on their functional role have been widely documented. However, their evolutionary dynamics and origin are still poorly understood. A general assumption is that intraspecific variation in a monomer of different satellite families can be fixed permanently in the genome [87]. Phylogenetically close lineages share common satellite sequences derived from an ancestral genome. Any differential copy number or acquired polymorphism within this sequence as a representative of a distant lineage can result in interspecific divergence [201,225,249,250]. Simultaneous occurrence of intraspecific homogenization of particular satellite repeat groups and fixation of species-specific mutations, as well as satellite conversion, all contribute to formation of a new species-specific satellite sequence. This phenomenon can have an important impact on driving molecular-level speciation [36]. Although the basic evolutionary mechanism that causes molecular organization and diversity is unknown, a common perception is that concerted evolution can result in unequal relocation of satellite repeats within the same or different chromosomes through various mechanisms, such as unequal crossing over, rolling-circular replication, gene conversion, and TE-mediated transfer [90,251]. Such events may subsequently trigger new means of amplification, thus giving rise to the formation of novel arrays of satDNA [62,127,228,232]. Certain chromosome-specific human AS groups have been highlighted with a specific age gradient for each chromosomal locus, which suggests that new AS repeats expanded in centromeric regions during evolution [29,180,229]. Data on AS sequence evolution in NHP species are rare, but several recent papers explore the mechanisms of evolution of primate satellites.

Ruiz-Ruano et al. [252] proposed a hypothesis stating that satellite repeats arise through de novo duplication of a genomic segment. As a result, a novel tandem repeat is formed at a different genomic position. Several possible mechanisms may facilitate this phenomenon, such as single-strand slippage during DNA replication or reinsertion of replicated copies from extrachromosomal DNA. The newly born satellite repeat can be dispersed throughout the genome by different processes, such as transposition or insertion of replicated copies of extrachromosomal DNA. This can eventually lead to amplification of certain loci and expansion in sequence size of the satellite repeat mostly by unequal crossing over. According to this evolutionary model, a tandem monomer of 15 bp or less can arise in several genomic loci randomly; therefore, both microsatellite and short minisatellite repeats may arise via this evolutionary process. Short repeated monomers can subsequently form long arrays by various continuously occurring processes of divergence and duplication [253,254]. An example of such arrays is HORs, which might be evolutionary products of homogenization and duplication of shorter tandem units [87]. The above-mentioned hypothesis proposed by Ruiz-Ruano et al. [252] can be integrated with another proposal, termed the “Library hypothesis”, which offers insights into the distribution of satDNA families among lineages [87,252,255–258]. As described in this section, species that are phylogenetically close share a common set of preserved satellite sequences derived from their ancestor, and each sequence shows differential amplification among each species.

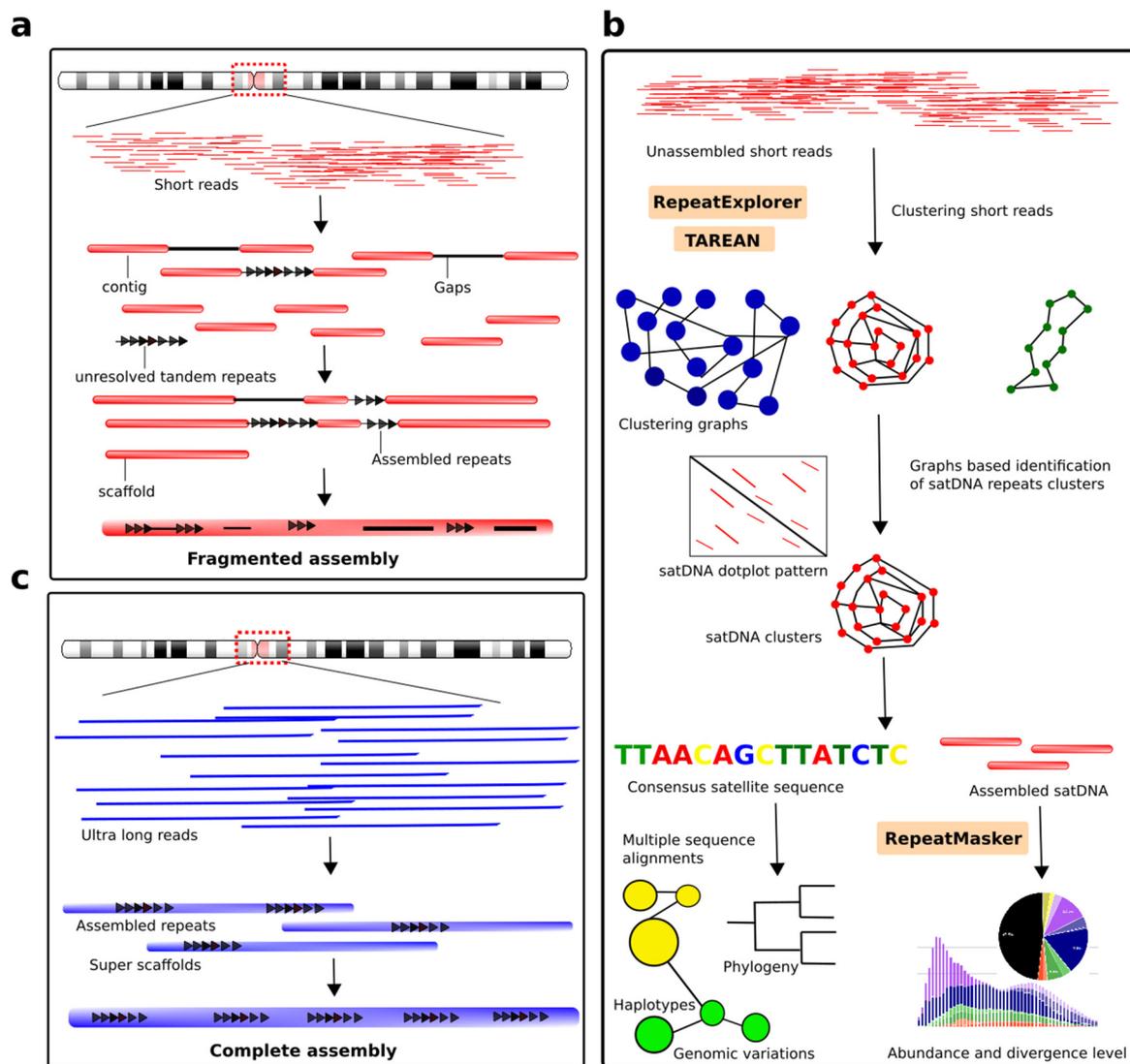
According to the Library hypothesis, the main driving forces of satellite repeat dissemination are transposition and rolling circle replication. Although no evidence has been reported that either of these mechanisms drives the spread of satDNA in the genome, the expansion of satellites has been suggested to occur through a rolling circular process [259,260]. There is increasing evidence to highlight that TEs can mediate the emergence of satellite sequences and their mobilization in the genome [258,261]. Transposable elements can be important contributors in shaping satDNA evolution by playing a crucial role in the formation of a new library of satellite repeats, their dispersion in the genome, and even their amplification into longer arrays in certain cases [258]. It is now considered that the origin and dissemination of new satellites mediated by transposition has been underestimated and there may be more cases of TEs and satDNA evolutionary association [262]. After the birth

and dissemination of a satDNA family within a genome, each satellite sequence might either evolve independently and freely through sequence divergence or follow concerted evolution. During this phase, all tandem repeated units may show cohesive evolution [263]. In contrast to what would likely happen in the absence of selective pressure, different units of a satDNA family could show a higher degree of intraspecific resemblance and interspecific divergence by following a concerted evolutionary process [36,87]. Gradual concerted evolution can occur in a step-wise manner through a specialized mechanism termed “molecular drive” [264,265]. This model proposes that new satDNA variants, formed through accumulation of mutations in the monomers, are distributed throughout the array by different mechanisms such as transposition, unequal crossing over, or reinsertion of extrachromosomal replicates (Figure 5). All these evolutionary events, together with gene conversion, can result in homogenization of evolved satellite sequences. These variants can then undergo fixation in the population by sexual reproduction. An important, interesting aspect of concerted evolution is the variable degree of population differentiation. Concerted evolution of satellite repeat sequences has been observed among populations in different species [227,250,263,266], which supports the hypothesis that satDNA might serve as the main driving force of population-level genomic variability.

## 8. Enlightening the Dark Matter of the Genome: Modern Approaches and Challenges in Detecting satDNA Repeats

Par-genome assemblies of model and non-model primate species are currently being produced at an impressive rate, with unprecedented quality and contiguity by consortia and individual laboratories [267]. However, difficulties in assembling repeat-rich regions (genomic “dark matter”) limits insights into the evolution of genome structure and regulatory networks [19]. The largest assembly gaps remain in centromeric regions and acrocentric short arms, sites known to contain megabase-sized arrays of satDNA [225]. Complex repeat structures have very important evolutionary and biomedical functions. Therefore, the successful assembly of these repeats completely and accurately is paramount to obtain maximum implications of primate genomes. The extremely repetitive nature of satDNA sequences has presented a challenging task to generate *de novo* genome assemblies. High levels of variation in the abundance of these tandem repeats among species cause complications in both assembly procedure and algorithms [44]. Ideally, a high-quality genome assembly must represent accurately annotated features such as genes and complete repeated contents with correct chromosomal/scaffold positions. The assembly can be further utilized in new experiments to provide a better understanding of gene or repeat function and expression, for instance during investigation of differential expression among diverse conditions using RNA sequencing [268]. Significant variation in the natural abundance of tandem repeats exists in different organisms. This complicates assembly procedures and prevents development of algorithms that perform reliably in all cases. The first-ever human repeatome database to include tandem repeats was introduced in 1992 [269] and is now developed as “Rebase” [270]. Subsequently, NGS has revolutionized the development of modern resources, including new approaches for detection of repeats and availability of repeat databases. More than 50 bioinformatics tools have been developed to detect tandem repeats, while numerous publicly accessible databases have been established containing an enormous amount of data applicable in various fields, such as medicine, agriculture, and forensics. Some well-known databases are the Human Genome Browser by UCSC [271], the Tandem Repeats Database (TRDB; [272], Dfam [273], and STRBase [274]. Most databases are based on data generated by well-known bioinformatic tools, e.g., “Tandem Repeats Finder” (TRF) [275] or “RepeatMasker” [52], which are widely used for the characterization of repeats. In particular, the RepeatMasker program is the preferred tool for repeat masking and identification in assembled genomes. However, the use of these two tools might not be sufficient to detect all repeats precisely because of their conservative approach, statistical errors, and lower prediction power in finding diverged repeats. To address these shortcomings, adequate statistical methodology integrated with a meta approach is required for accurate annotation of repeats. For instance, a program called “Tandem Repeat Annotation Library” (TRAL) [276] can

considerably improve repeats annotation, and further identify repeat regions that might be under selection. The repeat detectors, as described above, can be employed to annotate repeat contents including satDNA; however, these programs require a longer sequence (contigs) as input that can be sourced from genome assemblies. These assemblies are extremely under-represented for satDNA and can miss certain satellite sequences altogether (Figure 6a). This problem can be highly challenging in primate genome annotation because long reiterative arrays of HORs are hard to stitch together, particularly when they are composed of multiple mega-base sequences. Therefore, HORs and AS repeats were missing, and centromeric regions were recognized as assembly gaps in previous assembly versions of the human genome [55]. Graphical-reference models of HOR arrays were then constructed from whole-genome sequences to configure the HOR monomers and fill these gaps [55]. However, these models were not able to place the HOR units on chromosomes in a linear fashion nor could they combine the long-ranged units into a complete array. More recently, a highly contiguous human assembly was published using Nanopore ultra-long sequencing [277]. This assembly successfully stitched the linear array of AS sequences and the DYZ3 satellites that span the regions between the long and short arm of the Y chromosome [277]. This improvement in length of ultra-long Nanopore reads up to 1 Mb attracted global attention and enabled the development of new techniques to achieve assembly of AS arrays of other primates, making up the most abundant portion of their satellitomes (Figure 6b). It would be worthwhile to enhance the efficiency of detection of AS variants and their changes of copy number among different populations, species, or genera. The recent accomplishment of end-to-end assembly of the entire X chromosome of humans has further revolutionized genome science and opened the way to enable the detection of multi-megabase satellite arrays in the pericentromeric regions [8]. High coverage of ultra-long-read sequencing (PacBio and Oxford Nanopore) integrated with complementary technologies, such as Hi-C and BioNano, and development of additional methods are required to achieve gapless genomes and resolve the long stretches of centromeric satellites. However, generation of an accurate assembly is also a laborious and time-consuming task and requires deep sequencing, which can be costly, and in where the reference genome is unavailable, the approach will not be useful for characterizing satellite repeats. This could be a further challenge, especially when the goal of the study is to predict novel repeats. A similarity-based clustering algorithm has been developed that evaluates all sequence comparisons between unassembled reads and groups the repeated sequences into clusters [278] (Figure 6c). Two efficient pipelines based on this method are RepeatExplorer [279] and Tandem Repeat Analyzer (TAREAN) [280]. These tools have proved to be the most successful approaches to date, particularly in the characterization and quantification of satellite repeats.



**Figure 6.** Schematic illustration of genome assembly limitations and new next-generation sequencing approaches for satellite DNA analysis. (a) Centromere region of the human chromosome that could not be assembled using short-read sequencing and is not represented in primary human assemblies. The assembly algorithms cannot be used for short reads (red lines) of the centromere owing to high-level reiteration of tandem repeats (black triangles) and therefore could not be recovered in the assembly. The fragmented assembly may contain gaps, thereby missing satellite DNA sequences causing bias to genome annotation. (b) A cheap alternative to analyze the satellite directly is the development of clustering-based pipelines, which can graphically predict different repeats and cluster them into groups. These programs yield assembled contigs, which can be further utilized for downstream analyses, such as repeats abundance, divergence, and comparative genomic analysis. (c) Recent developments (ultra-long-read technology) have successfully recovered the complete human genome assembly [279].

## 9. Concluding Remarks

Satellite DNA is among the most fascinating components of the primate genome. Satellite repeats are extremely variable in sequence length. They have evolved rapidly and serve as vital sources for genetic divergence. Here, we have discussed the different aspects of satellite DNA evolutionary paradigms, genomic organization, diversity, and functional significance. We have reviewed different kinds of satDNA sequences that have been characterized in primate genomes, and outlined their proportion and structure as well as variation among primate lineages. These sequences are involved in different fundamental functions and are critically important in driving karyotypic evolution.

A remarkable feature of these elements is their intense degree of interspecific and intraspecific variation and divergence, which might be important factors in primate speciation. We have further reviewed hypotheses on the evolutionary origin, genomic birth, and expansion of satDNA. Presently, certain technological limitations hamper the progress of in-depth and precise analysis, which demands the development of efficient computational tools and sequencing technologies for complete assembly of the centromeric and telomeric regions to overcome these limitations and boost further studies. Significant efforts are required from the research community to elucidate these genomic regions, which have been ignored for too long.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4409/9/12/2714/s1>, Table S1: SatDNA repeats described in this review.

**Author Contributions:** S.F.A. and K.S. drafted the manuscript; S.F.A., W.S., M.J., and K.S. conceived the ideas and reviewed the literature; S.F.A. and M.J. performed the data analysis, including comparative repeat abundance and phylogeny, plotted graphics, and designed all figures; S.F.A., W.S., A.S., T.P., S.M., and K.S. reviewed the data and revised the manuscript. All authors approved the manuscript for publication. All authors have read and agreed to the published version of the manuscript.

**Funding:** This review was financially supported by grants from the Center for Advanced Studies in Tropical Natural Resources, National Research University-Kasetsart University (CASTNAR, NRU-KU, Thailand to K.S.), the Center for Advanced Studies in Tropical Natural Resources, National Research University-Kasetsart University (CASTNAR, NRU-KU, Thailand) awarded to K.S., the e-ASIA Joint Research Program (no. P1851131 to K.S. and W.S.), the National Science and Technology Development Agency (NSTDA) (NSTDA P-19-52238 to K.S. and W.S.), a Postdoctoral Researcher award at Kasetsart University (to S.F.A. and K.S.), the Thailand Research Fund (TRF) (Nos. RSA6180075 and PHD60I0014 to K.S. and W.S.), the Graduate Scholarship Program of the Graduate School, Kasetsart University, Thailand (to T.P. and K.S.), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES: 88882.433287/2019-01 to M.J.), the National Primate Research Center of Thailand-Chulalongkorn University (NPRCT-CU to K.S.), the Thailand Research Fund-Chinese Academy of Science (no. DBG608008 to S.M.), and Thailand Research Fund Senior Scholar (no. RTA6280010 to S.M.).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Rogers, J.; Gibbs, R.A. Comparative primate genomics: Emerging patterns of genome content and dynamics. *Nat. Rev. Genet.* **2014**, *15*, 347–359. [[CrossRef](#)] [[PubMed](#)]
2. Enard, W.; Pääbo, S. Comparative primate genomics. *Annu. Rev. Genom. Hum. Genet.* **2004**, *5*, 351–378. [[CrossRef](#)] [[PubMed](#)]
3. Mikkelsen, T.S.; Hillier, L.W.; Eichler, E.E.; Zody, M.C.; Jaffe, D.B.; Yang, S.P.; Enard, W.; Hellmann, I.; Lindblad-Toh, K.; Altheide, T.K.; et al. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **2005**, *437*, 69–87. [[CrossRef](#)]
4. Gibbs, R.A.; Rogers, J.; Katze, M.G.; Bumgarner, R.; Weinstock, G.M.; Mardis, E.R.; Remington, K.A.; Strausberg, R.L.; Venter, J.C.; Wilson, R.K.; et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **2007**, *316*, 222–234. [[CrossRef](#)] [[PubMed](#)]
5. Alföldi, J.; Lindblad-Toh, K. Comparative genomics as a tool to understand evolution and disease. *Genome Res.* **2013**, *23*, 1063–1068. [[CrossRef](#)]
6. Little, P.F.R. Structure and function of the human genome. *Genome Res.* **2005**, *15*, 1759–1765. [[CrossRef](#)]
7. Chi, K.R. The dark side of the human genome. *Nature* **2016**, *538*, 275–277. [[CrossRef](#)]
8. 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](#)]
9. Biémont, C. Genome size evolution: Within-species variation in genome size. *Heredity (Edinb)* **2008**, *101*, 297–298. [[CrossRef](#)]
10. Srikulnath, K.; Uno, Y.; Matsubara, K.; Thongpan, A.; Suputtitada, S.; Apisitwanich, S.; Nishida, C.; Matsuda, Y. Chromosomal localization of the 18S–28S and 5s rRNA genes and (TTAGGG)*n* sequences of butterfly lizards (*Leiolepis belliana belliana* and *Leiolepis boehmei*, Agamidae, Squamata). *Genet. Mol. Biol.* **2011**, *34*, 582–586. [[CrossRef](#)]

11. Ambrožová, K.; Mandáková, T.; Bureš, P.; Neumann, P.; Leitch, I.J.; Koblížková, A.; Macas, J.; Lysak, M.A. Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria lilies*. *Ann. Bot.* **2011**, *107*, 255–268. [[CrossRef](#)] [[PubMed](#)]
12. Scalvenzi, T.; Pollet, N. Insights on genome size evolution from a miniature inverted repeat transposon driving a satellite DNA. *Mol. Phylogenet. Evol.* **2014**, *81*, 1–9. [[CrossRef](#)] [[PubMed](#)]
13. Ahmad, S.F.; Singchat, W.; Jehangir, M.; Panthum, T.; Srikulnath, K. Consequence of paradigm shift with repeat landscapes in reptiles: Powerful facilitators of chromosomal rearrangements for diversity and evolution (running title: Genomic impact of repeats on chromosomal dynamics in reptiles). *Genes* **2020**, *11*, 827. [[CrossRef](#)] [[PubMed](#)]
14. Gregory, T.R.; Nicol, J.A.; Tamm, H.; Kullman, B.; Kullman, K.; Leitch, I.J.; Murray, B.G.; Kapraun, D.F.; Greilhuber, J.; Bennett, M.D. Eukaryotic genome size databases. *Nucleic Acids Res.* **2007**, *35*, D332–D338. [[CrossRef](#)]
15. Kidwell, M.G. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **2002**, *115*, 49–63. [[CrossRef](#)]
16. Hancock, J.M. Genome size and the accumulation of simple sequence repeats: Implications of new data from genome sequencing projects. *Genetica* **2002**, *115*, 93–103. [[CrossRef](#)]
17. Liu, G.; Thomas, J.; Touchman, J.; Blakesley, B.; Bouffard, G.; Beckstrom-Sternberg, S.; McDowell, J.; Maskeri, B.; Thomas, P.; Zhao, S.; et al. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* **2003**, *13*, 358–368. [[CrossRef](#)]
18. Marques-Bonet, T.; Ryder, O.A.; Eichler, E.E. Sequencing primate genomes: What have we learned? *Annu. Rev. Genomics Hum. Genet.* **2009**, *10*, 355–386. [[CrossRef](#)]
19. Treangen, T.J.; Salzberg, S.L. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nat. Rev. Genet.* **2012**, *13*, 36–46. [[CrossRef](#)]
20. Melters, D.P.; Bradnam, K.R.; Young, H.A.; Telis, N.; May, M.R.; Ruby, J.G.; Sebra, R.; Peluso, P.; Eid, J.; Rank, D.; et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **2013**, *14*, R10. [[CrossRef](#)]
21. Charlesworth, B.; Jarne, P.; Assimakopoulos, S. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. III. Element abundances in heterochromatin. *Genet. Res.* **1994**, *64*, 183–197. [[CrossRef](#)] [[PubMed](#)]
22. Prakhongcheep, O.; Hirai, Y.; Hara, T.; Srikulnath, K.; Hirai, H.; Koga, A. Two types of alpha satellite DNA in distinct chromosomal locations in Azara's owl monkey. *DNA Res.* **2013**, *20*, 235–240. [[CrossRef](#)] [[PubMed](#)]
23. Prakhongcheep, O.; Chaiprasertsri, N.; Terada, S.; Hirai, Y.; Srikulnath, K.; Hirai, H.; Koga, A. Heterochromatin blocks constituting the entire short arms of acrocentric chromosomes of Azara's owl monkey: Formation processes inferred from chromosomal locations. *DNA Res.* **2013**, *20*, 461–470. [[CrossRef](#)] [[PubMed](#)]
24. Prakhongcheep, O.; Thapana, W.; Suntronpong, A.; Singchat, W.; Pattanatanang, K.; Phatcharakullawarawat, R.; Muangmai, N.; Peyachoknagul, S.; Matsubara, K.; Ezaz, T.; et al. Lack of satellite DNA species-specific homogenization and relationship to chromosomal rearrangements in monitor lizards (Varanidae, Squamata). *BMC Evol. Biol.* **2017**, *17*, 193. [[CrossRef](#)]
25. Thongchum, R.; Singchat, W.; Laopichienpong, N.; Tawichasri, P.; Kraichak, E.; Prakhongcheep, O.; Sillapaprayoon, S.; Muangmai, N.; Baicharoen, S.; Suntrarachun, S.; et al. Diversity of PBI-DdeI satellite DNA in snakes correlates with rapid independent evolution and different functional roles. *Sci. Rep.* **2019**, *9*, 15459. [[CrossRef](#)]
26. Suntronpong, A.; Singchat, W.; Kruasuwan, W.; Prakhongcheep, O.; Sillapaprayoon, S.; Muangmai, N.; Somyong, S.; Indananda, C.; Kraichak, E.; Peyachoknagul, S.; et al. Characterization of centromeric satellite DNAs (MALREP) in the Asian swamp eel (*Monopterus albus*) suggests the possible origin of repeats from transposable elements. *Genomics* **2020**, *112*, 3097–3107. [[CrossRef](#)]
27. Nakagawa, T.; Okita, A.K. Transcriptional silencing of centromere repeats by heterochromatin safeguards chromosome integrity. *Curr. Genet.* **2019**, *65*, 1089–1098. [[CrossRef](#)]
28. Kim, J.H.; Ebersole, T.; Kouprina, N.; Noskov, V.N.; Ohzeki, J.I.; Masumoto, H.; Mravinac, B.; Sullivan, B.A.; Pavlicek, A.; Dovat, S.; et al. Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res.* **2009**, *19*, 533–544. [[CrossRef](#)]
29. Schueler, M.G.; Higgins, A.W.; Rudd, M.K.; Gustashaw, K.; Willard, H.F. Genomic and genetic definition of a functional human centromere. *Science* **2001**, *294*, 109–115. [[CrossRef](#)]

30. Schueler, M.G.; Sullivan, B.A. Structural and functional dynamics of human centromeric chromatin. *Annu. Rev. Genom. Hum. Genet.* **2006**, *7*, 301–313. [CrossRef]
31. Aldrup-MacDonald, M.E.; Sullivan, B.A. The past, present, and future of human centromere genomics. *Genes* **2014**, *5*, 33–50. [CrossRef] [PubMed]
32. Fachinetti, D.; Han, J.S.; McMahon, M.A.; Ly, P.; Abdullah, A.; Wong, A.J.; Cleveland, D.W. DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. *Dev. Cell* **2015**, *33*, 314–327. [CrossRef] [PubMed]
33. McNulty, S.M.; Sullivan, B.A. Alpha satellite DNA biology: Finding function in the recesses of the genome. *Chromosom. Res.* **2018**, *26*, 115–138. [CrossRef]
34. Jagannathan, M.; Warsinger-Pepe, N.; Watase, G.J.; Yamashita, Y.M. Comparative analysis of satellite DNA in the drosophila melanogaster species complex. *G3 Genes Genomes Genet.* **2017**, *7*, 693–704. [CrossRef]
35. Lower, S.S.; McGurk, M.P.; Clark, A.G.; Barbash, D.A. Satellite DNA evolution: Old ideas, new approaches. *Curr. Opin. Genet. Dev.* **2018**, *49*, 70–78. [CrossRef] [PubMed]
36. Garrido-Ramos, M.A. Satellite DNA: An evolving topic. *Genes* **2017**, *8*, 230. [CrossRef]
37. Hartley, G.; O’neill, R.J. Centromere repeats: Hidden gems of the genome. *Genes* **2019**, *10*, 223. [CrossRef] [PubMed]
38. Hedges, D.J.; Callinan, P.A.; Cordaux, R.; Xing, J.; Barnes, E.; Batzer, M.A. Differential Alu mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* **2004**, *14*, 1068–1075. [CrossRef]
39. Han, K.; Sen, S.K.; Wang, J.; Callinan, P.A.; Lee, J.; Cordaux, R.; Liang, P.; Batzer, M.A. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res.* **2005**, *33*, 4040–4052. [CrossRef]
40. Cordaux, R.; Batzer, M.A. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **2009**, *10*, 691–703. [CrossRef]
41. Cordaux, R.; Sen, S.K.; Konkel, M.K.; Batzer, M.A. Computational methods for the analysis of primate mobile elements. *Methods Mol. Biol.* **2010**, *628*, 137–151. [CrossRef] [PubMed]
42. Trizzino, M.; Park, Y.S.; Holsbach-Beltrame, M.; Aracena, K.; Mika, K.; Caliskan, M.; Perry, G.H.; Lynch, V.J.; Brown, C.D. Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* **2017**, *27*, 1623–1633. [CrossRef] [PubMed]
43. Jurka, J.; Krnjajic, M.; Kapitonov, V.V.; Stenger, J.E.; Kokhanyy, O. Active Alu in Paternal Germlines>Active Alu Elements Are Passed Primarily through Paternal Germlines. *Theor. Popul. Biol.* **2002**, *61*, 519–530. [CrossRef]
44. Tørresen, O.K.; Star, B.; Mier, P.; Andrade-Navarro, M.A.; Bateman, A.; Jarnot, P.; Gruca, A.; Grynberg, M.; Kajava, A.V.; Promponas, V.J.; et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res.* **2019**, *47*, 10994–11006. [CrossRef]
45. Lee, H.E.; Eo, J.; Kim, H.S. Composition and evolutionary importance of transposable elements in humans and primates. *Genes Genom.* **2014**, *37*, 135–140. [CrossRef]
46. Lee, H.R.; Hayden, K.E.; Willard, H.F. Organization and molecular evolution of CENP-A-associated satellite DNA families in a basal primate genome. *Genome Biol. Evol.* **2011**, *3*, 1136–1149. [CrossRef]
47. Cechova, M.; Harris, R.S.; Tomaszewicz, M.; Arbeithuber, B.; Chiaromonte, F.; Makova, K.D. High Satellite Repeat Turnover in Great Apes Studied with Short- And Long-Read Technologies. *Mol. Biol. Evol.* **2019**, *36*, 2415–2431. [CrossRef]
48. Cacheux, L.; Ponger, L.; Gerbault-Seureau, M.; Richard, F.A.; Escudé, C. Diversity and distribution of alpha satellite DNA in the genome of an Old World monkey: *Cercopithecus solatus*. *BMC Genom.* **2016**, *17*, 916. [CrossRef]
49. Cacheux, L.; Ponger, L.; Gerbault-Seureau, M.; Loll, F.; Gey, D.; Richard, F.A.; Escudé, C. The targeted sequencing of alpha satellite DNA in *Cercopithecus pogonias* provides new insight into the diversity and dynamics of centromeric repeats in old world monkeys. *Genome Biol. Evol.* **2018**, *10*, 1837–1851. [CrossRef]
50. Sullivan, L.L.; Chew, K.; Sullivan, B.A.  $\alpha$  satellite DNA variation and function of the human centromere. *Nucleus* **2017**, *8*, 331–339. [CrossRef]
51. Sullivan, L.L.; Sullivan, B.A. Genomic and functional variation of human centromeres. *Exp. Cell Res.* **2020**, *389*, 111896. [CrossRef]
52. Smit, A.; Hubley, R.; Grenn, P. RepeatMasker Open-4.0. 2015. Available online: <http://www.repeatmasker.org/> (accessed on 1 August 2020).

53. 10KTrees Website. Available online: <https://10ktrees.nunn-lab.org/Primates/dataset.html> (accessed on 27 July 2020).
54. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [[CrossRef](#)]
55. Miga, K.H. Completing the human genome: The progress and challenge of satellite DNA assembly. *Chromosom. Res.* **2015**, *23*, 421–426. [[CrossRef](#)] [[PubMed](#)]
56. Waring, M.; Britten, R.J. Nucleotide sequence repetition: A rapidly reassociating fraction of mouse DNA. *Science* **1966**, *154*, 791–794. [[CrossRef](#)] [[PubMed](#)]
57. López-Flores, I.; Garrido-Ramos, M.A. The repetitive DNA content of eukaryotic genomes. *Genome Dyn.* **2012**, *7*, 1–28. [[CrossRef](#)] [[PubMed](#)]
58. Biscotti, M.A.; Canapa, A.; Forconi, M.; Olmo, E.; Barucca, M. Transcription of tandemly repetitive DNA: Functional roles. *Chromosom. Res.* **2015**, *23*, 463–477. [[CrossRef](#)] [[PubMed](#)]
59. Rogers, J.; Mahaney, M.C.; Witte, S.M.; Nair, S.; Newman, D.; Wedel, S.; Rodriguez, L.A.; Rice, K.S.; Slifer, S.H.; Perelygin, A.; et al. A genetic linkage map of the baboon (*Papio hamadryas*) genome based on human microsatellite polymorphisms. *Genomics* **2000**, *67*, 237–247. [[CrossRef](#)]
60. Catacchio, C.R.; Ragone, R.; Chiatante, G.; Ventura, M. Organization and evolution of Gorilla centromeric DNA from old strategies to new approaches. *Sci. Rep.* **2015**, *5*, 14189. [[CrossRef](#)]
61. Bersani, F.; Lee, E.; Kharchenko, P.V.; Xu, A.W.; Liu, M.; Xega, K.; MacKenzie, O.C.; Brannigan, B.W.; Wittner, B.S.; Jung, H.; et al. Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15148–15153. [[CrossRef](#)]
62. Miga, K.H.; Newton, Y.; Jain, M.; Altemose, N.; Willard, H.F.; Kent, E.J. Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Res.* **2014**, *24*, 697–707. [[CrossRef](#)]
63. Sujjwattanasarat, P.; Thapana, W.; Srikulnath, K.; Hirai, Y.; Hirai, H.; Koga, A. Higher-order repeat structure in alpha satellite DNA occurs in New World monkeys and is not confined to hominoids. *Sci. Rep.* **2015**, *5*, 10315. [[CrossRef](#)] [[PubMed](#)]
64. Richard, G.F.; Pâques, F. Mini- and microsatellite expansions: The recombination connection. *EMBO Rep.* **2000**, *1*, 122–126. [[CrossRef](#)] [[PubMed](#)]
65. Subramanian, S.; Mishra, R.K.; Singh, L. Genome-wide analysis of microsatellite repeats in humans: Their abundance and density in specific genomic regions. *Genome Biol.* **2003**, *4*, R13. [[CrossRef](#)] [[PubMed](#)]
66. Ramel, C. Mini- and microsatellites. *EHP* **1997**, *105*, 781–789. [[CrossRef](#)] [[PubMed](#)]
67. Näslund, K.; Saetre, P.; Von Salomé, J.; Bergström, T.F.; Jareborg, N.; Jazin, E. Genome-wide prediction of human VNTRs. *Genomics* **2005**, *85*, 24–35. [[CrossRef](#)] [[PubMed](#)]
68. Blanquer-Maumont, A.; Crouau-Roy, B. Polymorphism, monomorphism, and sequences in conserved microsatellites in primate species. *J. Mol. Evol.* **1995**, *41*, 492–497. [[CrossRef](#)] [[PubMed](#)]
69. Garza, J.C.; Slatkin, M.; Freimer, N.B. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.* **1995**, *12*, 594–603. [[CrossRef](#)]
70. Coote, T.; Bruford, M.W. Human Microsatellites Applicable for Analysis of Genetic Variation in Apes and Old World Monkeys. *J. Hered.* **1996**, *87*, 406–410. [[CrossRef](#)]
71. Kayser, M.; Caglià, A.; Corach, D.; Fretwell, N.; Gehrig, C.; Graziosi, G.; Heidorn, F.; Herrmann, S.; Herzog, B.; Hidding, M.; et al. Evaluation of Y-chromosomal STRs: A multicenter study. *Int. J. Legal Med.* **1997**, *110*, 125–133. [[CrossRef](#)]
72. Goossens, B.; Chikhi, L.; Utami, S.S.; De Ruiter, J.; Bruford, M.W. A multi-samples, multi-extracts approach for microsatellite analysis of faecal samples in an arboreal ape. *Conserv. Genet.* **2000**, *1*, 157–162. [[CrossRef](#)]
73. Nair, S.; Ha, J.; Rogers, J. Nineteen new microsatellite DNA polymorphisms in pigtailed macaques (*Macaca nemestrina*). *Primates* **2000**, *41*, 343–350. [[CrossRef](#)] [[PubMed](#)]
74. Winkler, L.A.; Zhang, X.; Ferrell, R.; Wagner, R.; Dahl, J.; Peter, G.; Sohn, R. Geographic Microsatellite Variability in Central American Howling Monkeys. *Int. J. Primatol.* **2004**, *25*, 197–210. [[CrossRef](#)]
75. Clisson, I.; Lathuilliere, M.; Crouau-Roy, B. Conservation and evolution of microsatellite loci in primate taxa. *Am. J. Primatol.* **2000**, *50*, 205–214. [[CrossRef](#)]
76. Buschiazzi, E.; Gemmell, N.J. Conservation of human microsatellites across 450 million years of evolution. *Genome Biol. Evol.* **2010**, *2*, 153–165. [[CrossRef](#)]
77. Oklander, L.I.; Steinberg, E.R.; Mudry, M.D. A new world monkey microsatellite (AP74) highly conserved in primates. *Acta Biol. Colomb.* **2012**, *17*, 93–101.

78. Boán, F.; Blanco, M.G.; Quinteiro, J.; Mouriño, S.; Gómez-Márquez, J. Birth and Evolutionary History of a Human Minisatellite. *Mol. Biol. Evol.* **2004**, *21*, 228–235. [[CrossRef](#)]
79. Moyzis, R.K.; Buckingham, J.M.; Cram, L.S.; Dani, M.; Deaven, L.L.; Jones, M.D.; Meyne, J.; Ratliff, R.L.; Wu, J.R. A highly conserved repetitive DNA sequence, (TTAGGG)(n), present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6622–6626. [[CrossRef](#)]
80. O’Sullivan, R.J.; Karlseder, J. Telomeres: Protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 171–181. [[CrossRef](#)]
81. Bandaria, J.N.; Qin, P.; Berk, V.; Chu, S.; Yildiz, A. Shelterin protects chromosome ends by compacting telomeric chromatin. *Cell* **2016**, *164*, 735–746. [[CrossRef](#)]
82. Wyatt, H.D.M.; West, S.C.; Beattie, T.L. InTERTpreting telomerase structure and function. *Nucleic Acids Res.* **2010**, *38*, 5609–5622. [[CrossRef](#)]
83. Maddar, H.; Ratzkovsky, N.; Krauskopf, A. Role for telomere cap structure in meiosis. *Mol. Biol. Cell* **2001**, *12*, 3191–3203. [[CrossRef](#)] [[PubMed](#)]
84. Boán, F.; Rodríguez, J.M.; Gómez-Márquez, J. A non-hypervariable human minisatellite strongly stimulates in vitro intramolecular homologous recombination. *J. Mol. Biol.* **1998**, *278*, 499–505. [[CrossRef](#)] [[PubMed](#)]
85. Boán, F.; Rodríguez, J.M.; Mouriño, S.; Blanco, M.G.; Viñas, A.; Sánchez, L.; Gómez-Márquez, J. Recombination analysis of the human minisatellite MsH42 suggests the existence of two distinct pathways for initiation and resolution of recombination at MsH42 in rat testes nuclear extracts. *Biochemistry* **2002**, *41*, 2166–2176. [[CrossRef](#)] [[PubMed](#)]
86. Nergadze, S.G.; Rocchi, M.; Azzalin, C.M.; Mondello, C.; Giulotto, E. Insertion of telomeric repeats at intrachromosomal break sites during primate evolution. *Genome Res.* **2004**, *14*, 1704–1710. [[CrossRef](#)]
87. Plohl, M.; Meštrović, N.; Mravinac, B. Satellite DNA evolution. *Genome Dyn.* **2012**, *7*, 126–152. [[CrossRef](#)]
88. Kazakov, A.E.; Shepelev, V.A.; Tumeneva, I.G.; Alexandrov, A.A.; Yurov, Y.B.; Alexandrov, I.A. Interspersed repeats are found predominantly in the “old”  $\alpha$  satellite families. *Genomics* **2003**, *82*, 619–627. [[CrossRef](#)]
89. Steiner, F.A.; Henikoff, S. Diversity in the organization of centromeric chromatin. *Curr. Opin. Genet. Dev.* **2015**, *31*, 28–35. [[CrossRef](#)]
90. Plohl, M.; Luchetti, A.; Meštrović, N.; Mantovani, B. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* **2008**, *409*, 72–82. [[CrossRef](#)]
91. Verdaasdonk, J.S.; Bloom, K. Centromeres: Unique chromatin structures that drive chromosome segregation. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 320–332. [[CrossRef](#)]
92. Fukagawa, T.; Earnshaw, W.C. The centromere: Chromatin foundation for the kinetochore machinery. *Dev. Cell* **2014**, *30*, 496–508. [[CrossRef](#)]
93. Maio, J.J. DNA strand reassociation and polyribonucleotide binding in the African green monkey, *Cercopithecus aethiops*. *J. Mol. Biol.* **1971**, *56*, 579–595. [[CrossRef](#)]
94. Manuelidis, L.; Wu, J.C. Homology between human and simian repeated DNA. *Nature* **1978**, *276*, 92–94. [[CrossRef](#)] [[PubMed](#)]
95. Vissel, B.; Andy Choo, K.H. Evolutionary relationships of multiple alpha satellite subfamilies in the centromeres of human chromosomes 13, 14, and 21. *J. Mol. Evol.* **1992**, *35*, 137–146. [[CrossRef](#)] [[PubMed](#)]
96. Musich, P.R.; Brown, F.L.; Maio, J.J. Highly repetitive component  $\alpha$  and related alphoid DNAs in man and monkeys. *Chromosoma* **1980**, *80*, 331–348. [[CrossRef](#)] [[PubMed](#)]
97. Willard, H.F.; Waye, J.S. Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet.* **1987**, *3*, 192–198. [[CrossRef](#)]
98. Alves, G.; Seuánez, H.N.; Fanning, T. Alpha satellite DNA in neotropical primates (Platyrrhini). *Chromosoma* **1994**, *103*, 262–267. [[CrossRef](#)]
99. Alves, G.; Canavez, F.; Seuánez, H.; Fanning, T. Recently amplified satellite DNA in *Callithrix argentata* (Primates, Platyrrhini). *Chromosom. Res.* **1995**, *3*, 207–213. [[CrossRef](#)]
100. 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](#)]
101. Cellamare, A.; Catacchio, C.R.; Alkan, C.; Giannuzzi, G.; Antonacci, F.; Cardone, M.F.; Della Valle, G.; Malig, M.; Rocchi, M.; Eichler, E.E.; et al. New insights into centromere organization and evolution from the white-cheeked Gibbon and marmoset. *Mol. Biol. Evol.* **2009**, *26*, 1889–1900. [[CrossRef](#)]

102. Akihiko, K.; Yuriko, H.; Shoko, T.; Israt, J.; Sudarath, B.; Visit, A.; Hirohisa, H. Evolutionary origin of higher-order repeat structure in alpha-satellite DNA of primate centromeres. *DNA Res.* **2014**, *21*, 407–415. [[CrossRef](#)]
103. Plohl, M.; Meštrović, N.; Mravinac, B. Centromere identity from the DNA point of view. *Chromosoma* **2014**, *123*, 313–325. [[CrossRef](#)] [[PubMed](#)]
104. Pita, M.; Gosálvez, J.; Gosálvez, A.; Nieddu, M.; López-Fernández, C.; Mezzanotte, R. A highly conserved pericentromeric domain in human and gorilla chromosomes. *Cytogenet. Genome Res.* **2010**, *126*, 253–258. [[CrossRef](#)] [[PubMed](#)]
105. Jarmuz, M.; Glotzbach, C.D.; Bailey, K.A.; Bandyopadhyay, R.; Shaffer, L.G. The evolution of satellite III DNA subfamilies among primates. *Am. J. Hum. Genet.* **2007**, *80*, 495–501. [[CrossRef](#)] [[PubMed](#)]
106. 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](#)]
107. Miller, D.A. Evolution of primate chromosomes. *Science* **1977**, *198*, 1116–1124. [[CrossRef](#)] [[PubMed](#)]
108. Montefalcone, G.; Tempesta, S.; Rocchi, M.; Archidiacono, N. Centromere repositioning. *Genome Res.* **1999**, *9*, 1184–1188. [[CrossRef](#)]
109. Ventura, M.; Antonacci, F.; Cardone, M.F.; Stanyon, R.; D’Addabbo, P.; Cellamare, A.; Sprague, L.J.; Eichler, E.E.; Archidiacono, N.; Rocchi, M. Evolutionary formation of new centromeres in macaque. *Science* **2007**, *316*, 243–246. [[CrossRef](#)]
110. Stanyon, R.; Rocchi, M.; Capozzi, O.; Roberto, R.; Misceo, D.; Ventura, M.; Cardone, M.F.; Bigoni, F.; Archidiacono, N. Primate chromosome evolution: Ancestral karyotypes, marker order and neocentromeres. *Chromosom. Res.* **2008**, *16*, 17–39. [[CrossRef](#)]
111. Amor, D.J.; Andy Choo, K.H. Neocentromeres: Role in human disease, evolution, and centromere study. *Am. J. Hum. Genet.* **2002**, *71*, 695–714. [[CrossRef](#)]
112. Wade, C.M.; Giulotto, E.; Sigurdsson, S.; Zoli, M.; Gnerre, S.; Imsland, F.; Lear, T.L.; Adelson, D.L.; Bailey, E.; Bellone, R.R.; et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **2009**, *326*, 865–867. [[CrossRef](#)]
113. Shang, W.H.; Hori, T.; Toyoda, A.; Kato, J.; Pependorf, K.; Sakakibara, Y.; Fujiyama, A.; Fukagawa, T. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res.* **2010**, *20*, 1219–1228. [[CrossRef](#)]
114. Maio, J.J.; Brown, F.L.; Musich, P.R. Toward a molecular paleontology of primate genomes—I. The HindIII and EcoRI dimer families of alphoid DNAs. *Chromosoma* **1981**, *83*, 103–125. [[CrossRef](#)]
115. Kalitsis, P.; Choo, K.H.A. The evolutionary life cycle of the resilient centromere. *Chromosoma* **2012**, *121*, 327–340. [[CrossRef](#)] [[PubMed](#)]
116. McKinley, K.L.; Cheeseman, I.M. The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 16–29. [[CrossRef](#)] [[PubMed](#)]
117. Lee, J.; Hong, W.Y.; Cho, M.; Sim, M.; Lee, D.; Ko, Y.; Kim, J. Synteny Portal: A web-based application portal for synteny block analysis. *Nucleic Acids Res.* **2016**, *44*, W35–W40. [[CrossRef](#)] [[PubMed](#)]
118. Schneider, V.A.; Graves-Lindsay, T.; Howe, K.; Bouk, N.; Chen, H.C.; Kitts, P.A.; Murphy, T.D.; Pruitt, K.D.; Thibaud-Nissen, F.; Albracht, D.; et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **2017**, *27*, 849–864. [[CrossRef](#)] [[PubMed](#)]
119. Klein, S.J.; O’Neill, R.J. Transposable elements: Genome innovation, chromosome diversity, and centromere conflict. *Chromosom. Res.* **2018**, *26*, 5–23. [[CrossRef](#)]
120. Prosser, J.; Frommer, M.; Paul, C.; Vincent, P.C. Sequence relationships of three human satellite DNAs. *J. Mol. Biol.* **1986**, *187*, 145–155. [[CrossRef](#)]
121. Willard, H.F. Chromosome-specific organization of human alpha satellite DNA. *Am. J. Hum. Genet.* **1985**, *37*, 524–532.
122. Warburton, P.E.; Willard, H.F. Genomic analysis of sequence variation in tandemly repeated DNA. Evidence for localized homogeneous sequence domains within arrays of  $\alpha$ -satellite DNA. *J. Mol. Biol.* **1990**, *216*, 3–16. [[CrossRef](#)]
123. Paar, V.; Basar, I.; Rosandic, M.; Gluncic, M. Consensus Higher Order Repeats and Frequency of String Distributions in Human Genome. *Curr. Genom.* **2007**, *8*, 93–111. [[CrossRef](#)]

124. Aldrup-MacDonald, M.E.; Kuo, M.E.; Sullivan, L.L.; Chew, K.; Sullivan, B.A. Genomic variation within alpha satellite DNA influences centromere location on human chromosomes with metastable epialleles. *Genom. Res.* **2016**, *26*, 1301–1311. [[CrossRef](#)]
125. Willard, H.F. Evolution of alpha satellite. *Curr. Opin. Genet. Dev.* **1991**, *1*, 509–514. [[CrossRef](#)]
126. Haaf, T.; Warburton, P.E.; Willard, H.F. Integration of human  $\alpha$ -satellite DNA into simian chromosomes: Centromere protein binding and disruption of normal chromosome segregation. *Cell* **1992**, *70*, 681–696. [[CrossRef](#)]
127. 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](#)] [[PubMed](#)]
128. Haaf, T.; Willard, H.F. Chromosome-specific  $\alpha$ -satellite DNA from the centromere of chimpanzee chromosome 4. *Chromosoma* **1997**, *106*, 226–232. [[CrossRef](#)] [[PubMed](#)]
129. Terada, S.; Hirai, Y.; Hirai, H.; Koga, A. Higher-order repeat structure in alpha satellite DNA is an attribute of hominoids rather than hominids. *J. Hum. Genet.* **2013**, *58*, 752–754. [[CrossRef](#)]
130. Alkan, C.; Cardone, M.F.; Catacchio, C.R.; Antonacci, F.; O'Brien, S.J.; Ryder, O.A.; Purgato, S.; Zoli, M.; Della Valle, G.; Eichler, E.E.; et al. Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genom. Res.* **2011**, *21*, 137–145. [[CrossRef](#)]
131. Koga, A.; Tanabe, H.; Hirai, Y.; Imai, H.; Imamura, M.; Oishi, T.; Stanyon, R.; Hirai, H. Co-opted megasatellite DNA drives evolution of secondary night vision in Azara's Owl monkey. *Genome Biol. Evol.* **2017**, *9*, 1963–1970. [[CrossRef](#)]
132. Nishihara, H.; Stanyon, R.; Kusumi, J.; Hirai, H.; Koga, A. Evolutionary origin of OwlRep, a megasatellite DNA associated with adaptation of owl monkeys to nocturnal lifestyle. *Genome Biol. Evol.* **2018**, *10*, 157–165. [[CrossRef](#)]
133. Wayne, J.S.; Willard, H.F. Human  $\beta$  satellite DNA: Genomic organization and sequence definition of a class of highly repetitive tandem DNA. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 6250–6254. [[CrossRef](#)] [[PubMed](#)]
134. 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](#)]
135. Cardone, M.F.; Ballarati, L.; Ventura, M.; Rocchi, M.; Marozzi, A.; Ginelli, E.; Meneveri, R. Evolution of beta satellite DNA sequences: Evidence for duplication-mediated repeat amplification and spreading. *Mol. Biol. Evol.* **2004**, *21*, 1792–1799. [[CrossRef](#)] [[PubMed](#)]
136. Meneveri, R.; Agresti, A.; Valle, G.D.; Talarico, D.; Siccardi, A.G.; Ginelli, E. Identification of a human clustered G + C-rich DNA family of repeats (Sau3A family). *J. Mol. Biol.* **1985**, *186*, 483–489. [[CrossRef](#)]
137. Meneveri, R.; Agresti, A.; Marozzi, A.; Saccone, S.; Rocchi, M.; Archidiacono, N.; Corneo, G.; Valle, G.D.; Ginelli, E. Molecular organization and chromosomal location of human GC-rich heterochromatic blocks. *Gene* **1993**, *123*, 227–234. [[CrossRef](#)]
138. Agresti, A.; Rainaldi, G.; Lobbiani, A.; Magnani, I.; Di Lernia, R.; Meneveri, R.; Siccardi, A.G.; Ginelli, E. Chromosomal location by in situ hybridization of the human Sau3A family of DNA repeats. *Hum. Genet.* **1987**, *75*, 326–332. [[CrossRef](#)]
139. Agresti, A.; Meneveri, R.; Siccardi, A.G.; Marozzi, A.; Corneo, G.; Gaudi, S.; Ginelli, E. Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequences (HaeIII family). *J. Mol. Biol.* **1989**, *205*, 625–631. [[CrossRef](#)]
140. Bakker, E.; Wijmenga, C.; Vossen, R.H.A.M.; Padberg, G.W.; Hewitt, J.; van Der Wielen, M.; Rasmussen, K.; Frants, R.R. The FSHD-linked locus D4F104S1 (p13E-11) ON 4q35 has a homologue on 10qter. *Muscle Nerve* **1995**, *18*, S39–S44. [[CrossRef](#)]
141. Lemmers, R.J.F.L.; Wohlgemuth, M.; Frants, R.R.; Padberg, G.W.; Morava, E.; Van Der Maarel, S.M. Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **2004**, *75*, 1124–1130. [[CrossRef](#)]
142. Clark, L.N.; Koehler, U.; Ward, D.C.; Wienberg, J.; Hewitt, J.E. Analysis of the organisation and localisation of the FSHD-associated tandem array in primates: Implications for the origin and evolution of the 3.3 kb repeat family. *Chromosoma* **1996**, *105*, 180–189. [[CrossRef](#)]

143. Winokur, S.T.; Bengtsson, U.; Vargas, J.C.; Wasmuth, J.J.; Altherr, M.R. The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum. Mol. Genet.* **1996**, *5*, 1567–1575. [[CrossRef](#)] [[PubMed](#)]
144. Ballarati, L.; Piccini, I.; Carbone, L.; Archidiacono, N.; Rollier, A.; Marozzi, A.; Meneveri, R.; Ginelli, E. Human genome dispersal and evolution of 4q35 duplications and interspersed LSau repeats. *Gene* **2002**, *296*, 21–27. [[CrossRef](#)]
145. Meneveri, R.; Agresti, A.; Rocchi, M.; Marozzi, A.; Ginelli, E. Analysis of GC-rich repetitive nucleotide sequences in great apes. *J. Mol. Evol.* **1995**, *40*, 405–412. [[CrossRef](#)] [[PubMed](#)]
146. Hirai, H.; Taguchi, T.; Godwin, A.K. Genomic differentiation of 18S ribosomal DNA and  $\beta$ -satellite DNA in the hominoid and its evolutionary aspects. *Chromosom. Res.* **1999**, *7*, 531–540. [[CrossRef](#)] [[PubMed](#)]
147. McLaughlin, C.R.; Chadwick, B.P. Characterization of DXZ4 conservation in primates implies important functional roles for CTCF binding, array expression and tandem repeat organization on the X chromosome. *Genome Biol.* **2011**, *12*, R37. [[CrossRef](#)] [[PubMed](#)]
148. Mitchell, A.R.; Gosden, J.R.; Ryder, O.A. Satellite DNA relationships in man and the primates. *Nucleic Acids Res.* **1981**, *9*, 3235–3249. [[CrossRef](#)]
149. Fowler, J.C.S.; Burgoyne, L.A.; Baker, E.G.; Riugenbergs, M.L.; Callen, D.F. Human Satellite III DNA: Genomic location and sequence homogeneity of the TaqI-deficient polymorphic sequences. *Chromosoma* **1989**, *98*, 266–272. [[CrossRef](#)]
150. Hu, H.; Li, B.; Duan, S. The alteration of subtelomeric DNA methylation in aging-related diseases. *Front. Genet.* **2019**, *10*, 697. [[CrossRef](#)]
151. Vergnaud, G.; Denoed, F. Minisatellites: Mutability and genome architecture. *Genome Res.* **2000**, *10*, 899–907. [[CrossRef](#)]
152. Riethman, H. Human subtelomeric copy number variations. *Cytogenet. Genome Res.* **2009**, *123*, 244–252. [[CrossRef](#)]
153. Louis, E.J.; Vershinin, A.V. Chromosome ends: Different sequences may provide conserved functions. *BioEssays* **2005**, *27*, 685–697. [[CrossRef](#)]
154. Cuadrado, A.; Jouve, N. Mapping and organization of highly-repeated DNA sequences by means of simultaneous and sequential FISH and C-banding in 6 $\times$ -triticale. *Chromosom. Res.* **1994**, *2*, 331–338. [[CrossRef](#)] [[PubMed](#)]
155. Brown, W.R.A.; MacKinnon, P.J.; Villasanté, A.; Spurr, N.; Buckle, V.J.; Dobson, M.J. Structure and polymorphism of human telomere-associated DNA. *Cell* **1990**, *63*, 119–132. [[CrossRef](#)]
156. Jurka, J.; Pethiyagoda, C. Simple repetitive DNA sequences from primates: Compilation and analysis. *J. Mol. Evol.* **1995**, *40*, 120–126. [[CrossRef](#)]
157. Araujo, N.P.; De Lima, L.G.; Dias, G.B.; Kuhn, G.C.S.; De Melo, A.L.; Yonenaga-Yassuda, Y.; Stanyon, R.; Svartman, M. Identification and characterization of a subtelomeric satellite DNA in Callitrichini monkeys. *DNA Res.* **2017**, *24*, 377–385. [[CrossRef](#)] [[PubMed](#)]
158. Royle, N.J.; Baird, D.M.; Jeffreys, A.J. A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. *Nat. Genet.* **1994**, *6*, 52–56. [[CrossRef](#)] [[PubMed](#)]
159. Koga, A.; Hirai, Y.; Hara, T.; Hirai, H. Repetitive sequences originating from the centromere constitute large-scale heterochromatin in the telomere region in the siamang, a small ape. *Heredity (Edinb)* **2012**, *109*, 180–187. [[CrossRef](#)]
160. Ventura, M.; Catacchio, C.R.; Sajjadian, S.; Vives, L.; Sudmant, P.H.; Marques-Bonet, T.; Graves, T.A.; Wilson, R.K.; Eichler, E.E. The evolution of African great ape subtelomeric heterochromatin and the fusion of human chromosome 2. *Genome Res.* **2012**, *22*, 1036–1049. [[CrossRef](#)]
161. Fanning, T.G.; Seuánez, H.N.; Forman, L. Satellite DNA sequences in the New World primate *Cebus apella* (Platyrrhini, Primates). *Chromosoma* **1993**, *102*, 306–311. [[CrossRef](#)]
162. Nishibuchi, G.; Déjardin, J. The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. *Chromosom. Res.* **2017**, *25*, 77–87. [[CrossRef](#)]
163. Riethman, H.; Ambrosini, A.; Paul, S. Human subtelomere structure and variation. *Chromosom. Res.* **2005**, *13*, 505–515. [[CrossRef](#)] [[PubMed](#)]
164. Kehrer-Sawatzki, H.; Cooper, D.N. Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosom. Res.* **2008**, *16*, 41–56. [[CrossRef](#)] [[PubMed](#)]

165. Stankiewicz, P.; Park, S.S.; Inoue, K.; Lupski, J.R. The evolutionary chromosome translocation 4;19 in Gorilla gorilla is associated with microduplication of the chromosome fragment syntenic to sequences surrounding the human proximal CMT1A-REP. *Genome Res.* **2001**, *11*, 1205–1210. [[CrossRef](#)] [[PubMed](#)]
166. Carbone, L.; Alan Harris, R.; Gnerre, S.; Veeramah, K.R.; Lorente-Galdos, B.; Huddleston, J.; Meyer, T.J.; Herrero, J.; Roos, C.; Aken, B.; et al. Gibbon genome and the fast karyotype evolution of small apes. *Nature* **2014**, *513*, 195–201. [[CrossRef](#)] [[PubMed](#)]
167. Koehler, U.; Bigoni, F.; Wienberg, J.; Stanyon, R. Genomic reorganization in the concolor gibbon (*hylobates concolor*) revealed by chromosome painting. *Genomics* **1995**, *30*, 287–292. [[CrossRef](#)]
168. Wienberg, J.; Jauch, A.; Lüdecke, H.J.; Senger, G.; Horsthemke, B.; Claussen, U.; Cremer, T.; Arnold, N.; Lengauer, C. The origin of human chromosome 2 analyzed by comparative chromosome mapping with a DNA microlibrary. *Chromosom. Res.* **1994**, *2*, 405–410. [[CrossRef](#)]
169. Cheng, Z.; Ventura, M.; She, X.; Khaitovich, P.; Graves, T.; Osoegawa, K.; Church, D.; DeJong, P.; Wilson, R.K.; Pääbo, S.; et al. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **2005**, *437*, 88–93. [[CrossRef](#)]
170. Ahmad, S.; Martins, C. The Modern View of B Chromosomes under the Impact of High Scale Omics Analyses. *Cells* **2019**, *8*, 156. [[CrossRef](#)]
171. Singh, L.; Purdom, I.F.; Jones, K.W. Sex chromosome associated satellite DNA: Evolution and conservation. *Chromosoma* **1980**, *79*, 137–157. [[CrossRef](#)]
172. Hughes, J.F.; Rozen, S. Genomics and genetics of human and primate y chromosomes. *Annu. Rev. Genom. Hum. Genet.* **2012**, *13*, 83–108. [[CrossRef](#)]
173. Steinemann, S.; Steinemann, M. Y chromosomes: Born to be destroyed. *BioEssays* **2005**, *27*, 1076–1083. [[CrossRef](#)] [[PubMed](#)]
174. Filatov, D.A.; Monéger, F.; Negrutlu, I.; Charlesworth, D. Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **2000**, *404*, 388–390. [[CrossRef](#)] [[PubMed](#)]
175. Charlesworth, B.; Charlesworth, D. The degeneration of Y chromosomes. *Philos. Trans. R. Soc. B Biol. Sci.* **2000**, *355*, 1563–1572. [[CrossRef](#)] [[PubMed](#)]
176. Skaletsky, H.; Kuroda-Kawaguchi, T.; Minx, P.J.; Cordum, H.S.; Hillier, L.D.; Brown, L.G.; Reppng, S.; Pyntikova, T.; All, J.; Blerl, T.; et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **2003**, *423*, 825–837. [[CrossRef](#)] [[PubMed](#)]
177. Blackmon, H.; Brandvain, Y. Long-term fragility of Y chromosomes is dominated by short-term resolution of sexual antagonism. *Genetics* **2017**, *207*, 1621–1629. [[CrossRef](#)] [[PubMed](#)]
178. Lee, C.; Li, X.; Jabs, E.W.; Court, D.; Lin, C.C. Human gamma X satellite DNA: An X chromosome specific centromeric DNA sequence. *Chromosoma* **1995**, *104*, 103–112. [[CrossRef](#)]
179. Lee, C.; Stanyon, R.; Lin, C.C.; Ferguson-Smith, M.A. Conservation of human gamma-X centromeric satellite DNA among primates with an autosomal localization in certain Old World monkeys. *Chromosom. Res.* **1999**, *7*, 43–47. [[CrossRef](#)]
180. Schueler, M.G.; Dunn, J.M.; Bird, C.P.; Ross, M.T.; Viggiano, L.; Rocchi, M.; Willard, H.F.; Green, E.D. Progressive proximal expansion of the primate X chromosome centromere. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10563–10568. [[CrossRef](#)]
181. Babcock, M.; Yatsenko, S.; Stankiewicz, P.; Lupski, J.R.; Morrow, B.E. AT-rich repeats associated with chromosome 22q11.2 rearrangement disorders shape human genome architecture on Yq12. *Genome Res.* **2007**, *17*, 451–460. [[CrossRef](#)]
182. Hughes, J.F.; Skaletsky, H.; Pyntikova, T.; Graves, T.A.; Van Daalen, S.K.M.; Minx, P.J.; Fulton, R.S.; McGrath, S.D.; Locke, D.P.; Friedman, C.; et al. Chimpanzee and human y chromosomes are remarkably divergent in structure and gene content. *Nature* **2010**, *463*, 536–539. [[CrossRef](#)]
183. Tomaszewicz, M.; Rangavittal, S.; Cechova, M.; Sanchez, R.C.; Fescemyer, H.W.; Harris, R.; Ye, D.; O'Brien, P.C.M.; Chikhi, R.; Ryder, O.A.; et al. A time- and cost-effective strategy to sequence mammalian Y chromosomes: An application to the de novo assembly of gorilla Y. *Genome Res.* **2016**, *26*, 530–540. [[CrossRef](#)] [[PubMed](#)]
184. Bernardo Carvalho, A.; Clark, A.G. Efficient identification of y chromosome sequences in the human and drosophila genomes. *Genome Res.* **2013**, *23*, 1894–1907. [[CrossRef](#)] [[PubMed](#)]
185. Brutlag, D.L. Molecular arrangement and evolution of heterochromatic DNA. *Annu. Rev. Genet.* **1980**, *14*, 121–144. [[CrossRef](#)]

186. Nei, M. Accumulation of Nonfunctional Genes on Sheltered Chromosomes. *Am. Nat.* **1970**, *104*, 311–322. [[CrossRef](#)]
187. Bachtrog, D. Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **2013**, *14*, 113–124. [[CrossRef](#)] [[PubMed](#)]
188. Floridia, G.; Zatterale, A.; Zuffardi, O.; Tyler-Smith, C. Mapping of a human centromere onto the DNA by topoisomerase II cleavage. *EMBO Rep.* **2000**, *1*, 489–493. [[CrossRef](#)] [[PubMed](#)]
189. Andersen, C.L.; Wandall, A.; Kjeldsen, E.; Mielke, C.; Koch, J. Active, but not inactive, human centromeres display topoisomerase II activity in vivo. *Chromosom. Res.* **2002**, *10*, 305–312. [[CrossRef](#)]
190. Spence, J.M.; Critcher, R.; Ebersole, T.A.; Valdivia, M.M.; Earnshaw, W.C.; Fukagawa, T.; Farr, C.J. Co-localization of centromere activity, proteins and topoisomerase II within a subdomain of the major human X  $\alpha$ -satellite array. *EMBO J.* **2002**, *21*, 5269–5280. [[CrossRef](#)]
191. Giacalone, J.; Friedes, J.; Francke, U. A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. *Nat. Genet.* **1992**, *1*, 137–143. [[CrossRef](#)]
192. Samonte, R.V.; Conte, R.A.; Verma, R.S. Physical mapping of human 7q and 14q subtelomeric DNA sequences in the great apes. *DNA Res.* **1997**, *4*, 249–252. [[CrossRef](#)]
193. Graur, D.; Zheng, Y.; Azevedo, R.B.R. An evolutionary classification of genomic function. *Genome Biol. Evol.* **2015**, *7*, 642–645. [[CrossRef](#)]
194. Palacios-Gimenez, O.M.; Bardella, V.B.; Lemos, B.; Cabral-De-Mello, D.C. Satellite DNAs are conserved and differentially transcribed among *Gryllus* cricket species. *DNA Res.* **2018**, *25*, 137–147. [[CrossRef](#)] [[PubMed](#)]
195. Ferreira, D.; Meles, S.; Escudeiro, A.; Mendes-da-Silva, A.; Adegas, F.; Chaves, R. Satellite non-coding RNAs: The emerging players in cells, cellular pathways and cancer. *Chromosom. Res.* **2015**, *23*, 479–493. [[CrossRef](#)] [[PubMed](#)]
196. Louzada, S.; Lopes, M.; Ferreira, D.; Adegas, F.; Escudeiro, A.; Gama-carvalho, M.; Chaves, R. Decoding the role of satellite DNA in genome architecture and plasticity—An evolutionary and clinical affair. *Genes* **2020**, *11*, 72. [[CrossRef](#)] [[PubMed](#)]
197. Quénet, D.; Dalal, Y. A long non-coding RNA is required for targeting centromeric protein A to the human centromere. *eLife* **2014**, *3*, e26016. [[CrossRef](#)]
198. Johnson, W.L.; Yewdell, W.T.; Bell, J.C.; McNulty, S.M.; Duda, Z.; O'Neill, R.J.; Sullivan, B.A.; Straight, A.F. RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. *eLife* **2017**, *6*, e25299. [[CrossRef](#)] [[PubMed](#)]
199. Liu, H.; Qu, Q.; Warrington, R.; Rice, A.; Cheng, N.; Yu, H. Mitotic Transcription Installs Sgo1 at Centromeres to Coordinate Chromosome Segregation. *Mol. Cell* **2015**, *59*, 426–436. [[CrossRef](#)] [[PubMed](#)]
200. Camacho, O.V.; Galan, C.; Swist-Rosowska, K.; Ching, R.; Gamalinda, M.; Karabiber, F.; De La Rosa-Velazquez, I.; Engist, B.; Koschorz, B.; Shukeir, N.; et al. Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. *eLife* **2017**, *6*, e25293. [[CrossRef](#)]
201. McNulty, S.M.; Sullivan, L.L.; Sullivan, B.A. Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. *Dev. Cell* **2017**, *42*, 226–240.e6. [[CrossRef](#)]
202. Shirai, A.; Kawaguchi, T.; Shimojo, H.; Muramatsu, D.; Ishida-Yonetani, M.; Nishimura, Y.; Kimura, H.; Nakayama, J.I.; Shinkai, Y. Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. *eLife* **2017**, *6*, e25317. [[CrossRef](#)]
203. Chan, F.L.; Marshall, O.J.; Saffery, R.; Kim, B.W.; Earle, E.; Choo, K.H.A.; Wong, L.H. Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1979–1984. [[CrossRef](#)] [[PubMed](#)]
204. Lyn Chan, F.; Wong, L.H. Transcription in the maintenance of centromere chromatin identity. *Nucleic Acids Res.* **2012**, *40*, 11178–11188. [[CrossRef](#)] [[PubMed](#)]
205. Grenfell, A.W.; Heald, R.; Strzelecka, M. Mitotic noncoding RNA processing promotes kinetochore and spindle assembly in *Xenopus*. *J. Cell Biol.* **2016**, *214*, 133–141. [[CrossRef](#)] [[PubMed](#)]
206. Smurova, K.; De Wulf, P. Centromere and Pericentromere Transcription: Roles and Regulation . . . in *Sickness and in Health*. *Front. Genet.* **2018**, *9*, 674. [[CrossRef](#)]
207. Jolly, C.; Metz, A.; Govin, J.; Vigneron, M.; Turner, B.M.; Khochbin, S.; Vourc'h, C. Stress-induced transcription of satellite III repeats. *J. Cell Biol.* **2004**, *164*, 25–33. [[CrossRef](#)]

208. Rizzi, N.; Denegri, M.; Chiodi, I.; Corioni, M.; Valgardsdottir, R.; Cobiانchi, F.; Riva, S.; Biamonti, G. Transcriptional Activation of a Constitutive Heterochromatin Domain of the Human Genome in Response to Heat Shock. *Mol. Biol. Cell* **2004**, *15*, 543–551. [[CrossRef](#)]
209. Biamonti, G. Nuclear stress bodies: A heterochromatin affair? *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 493–498. [[CrossRef](#)]
210. Goenka, A.; Sengupta, S.; Pandey, R.; Parihar, R.; Mohanta, G.C.; Mukerji, M.; Ganesh, S. Human satellite-III non-coding RNAs modulate heat-shock-induced transcriptional repression. *J. Cell Sci.* **2016**, *129*, 3541–3552. [[CrossRef](#)]
211. Valgardsdottir, R.; Chiodi, I.; Giordano, M.; Rossi, A.; Bazzini, S.; Ghigna, C.; Riva, S.; Biamonti, G. Transcription of Satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Res.* **2008**, *36*, 423–434. [[CrossRef](#)]
212. Pezer, Ž.; Brajković, J.; Feliciello, I.; Ugarković, D. Satellite DNA-mediated effects on genome regulation. *Genome Dyn.* **2012**, *7*, 153–169. [[CrossRef](#)]
213. Lanza, R.P.; Cibelli, J.B.; Blackwell, C.; Cristofalo, V.J.; Francis, M.K.; Baerlocher, G.M.; Mak, J.; Schertzer, M.; Chavez, E.A.; Sawyer, N.; et al. Extension of cell life-span and telomere length in animals cloned from senescent somatic cells. *Science* **2000**, *288*, 665–669. [[CrossRef](#)] [[PubMed](#)]
214. Rizvi, S.; Raza, S.T.; Mahdi, F. Telomere Length Variations in Aging and Age-Related Diseases. *Curr. Aging Sci.* **2015**, *7*, 161–167. [[CrossRef](#)] [[PubMed](#)]
215. Bass, H.W.; Riera-Lizarazu, O.; Ananiev, E.V.; Bordoli, S.J.; Rines, H.W.; Phillips, R.L.; Sedat, J.W.; Agard, D.A.; Cande, W.Z. Evidence for the coincident initiation of homolog pairing and synapsis during the telomere-clustering (bouquet) stage of meiotic prophase. *J. Cell Sci.* **2000**, *113*, 1033–1042. [[PubMed](#)]
216. Koga, A.; Notohara, M.; Hirai, H. Evolution of subterminal satellite (StSat) repeats in hominids. *Genetica* **2011**, *139*, 167–175. [[CrossRef](#)] [[PubMed](#)]
217. Novo, C.; Arnoult, N.; Bordes, W.Y.; Castro-Vega, L.; Gibaud, A.; Dutrillaux, B.; Bacchetti, S.; Londoño-Vallejo, A. The heterochromatic chromosome caps in great apes impact telomere metabolism. *Nucleic Acids Res.* **2013**, *41*, 4792–4801. [[CrossRef](#)] [[PubMed](#)]
218. Calderón, M.D.C.; Rey, M.D.; Cabrera, A.; Prieto, P. The subtelomeric region is important for chromosome recognition and pairing during meiosis. *Sci. Rep.* **2014**, *4*, 6488. [[CrossRef](#)]
219. Eymery, A.; Horard, B.; el Atifi-Borel, M.; Fourel, G.; Berger, F.; Vitte, A.L.; Van den Broeck, A.; Brambilla, E.; Fournier, A.; Callanan, M.; et al. A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. *Nucleic Acids Res.* **2009**, *37*, 6340–6354. [[CrossRef](#)]
220. Ting, D.T.; Lipson, D.; Paul, S.; Brannigan, B.W.; Akhavanfard, S.; Coffman, E.J.; Contino, G.; Deshpande, V.; Iafrate, A.J.; Letovsky, S.; et al. Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* **2011**, *331*, 593–596. [[CrossRef](#)]
221. Ugarković, D.; Plohl, M. Variation in satellite DNA profiles—Causes and effects. *EMBO J.* **2002**, *21*, 5955–5959. [[CrossRef](#)]
222. Gläser, B.; Grützner, F.; Willmann, U.; Stanyon, R.; Arnold, N.; Taylor, K.; Rietschel, W.; Zeitler, S.; Toder, R.; Schempp, W. Simian Y Chromosomes: Species-specific rearrangements of DAZ, RBM, and TSPY versus contiguity of PAR and SRY. *Mamm. Genome* **1998**, *9*, 226–231. [[CrossRef](#)]
223. Altemose, N.; Miga, K.H.; Maggioni, M.; Willard, H.F. Genomic Characterization of Large Heterochromatic Gaps in the Human Genome Assembly. *PLoS Comput. Biol.* **2014**, *10*, e1003628. [[CrossRef](#)] [[PubMed](#)]
224. Alonso, A.; Hasson, D.; Cheung, F.; Warburton, P.E. A paucity of heterochromatin at functional human neocentromeres. *Epigenetics Chromatin* **2010**, *3*, 6. [[CrossRef](#)] [[PubMed](#)]
225. Miga, K.H. Centromeric satellite DNAs: Hidden sequence variation in the human population. *Genes* **2019**, *10*, 352. [[CrossRef](#)] [[PubMed](#)]
226. Wayne, 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. [[CrossRef](#)] [[PubMed](#)]
227. Warburton, P.E.; Willard, H.F. Interhomologue sequence variation of alpha satellite DNA from human chromosome 17: Evidence for concerted evolution along haplotypic lineages. *J. Mol. Evol.* **1995**, *41*, 1006–1015. [[CrossRef](#)] [[PubMed](#)]
228. Rudd, M.K.; Wray, G.A.; Willard, H.F. The evolutionary dynamics of  $\alpha$ -satellite. *Genome Res.* **2006**, *16*, 88–96. [[CrossRef](#)] [[PubMed](#)]

229. 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](#)]
230. Maloney, K.A.; Sullivan, L.L.; Matheny, J.E.; Strome, E.D.; Merrett, S.L.; Ferris, A.; Sullivan, B.A. Functional epialleles at an endogenous human centromere. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 13704–13709. [[CrossRef](#)]
231. Durfy, S.J.; Willard, H.F. Molecular analysis of a polymorphic domain of alpha satellite from the human X chromosome. *Am. J. Hum. Genet.* **1987**, *41*, 391–401.
232. Schindelhauer, D.; Schwarz, T. Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous  $\alpha$ -satellite DNA array. *Genome Res.* **2002**, *12*, 1815–1826. [[CrossRef](#)]
233. Goldberg, I.G.; Sawhney, H.; Pluta, A.F.; Warburton, P.E.; Earnshaw, W.C. Surprising deficiency of CENP-B binding sites in African green monkey alpha-satellite DNA: Implications for CENP-B function at centromeres. *Mol. Cell. Biol.* **1996**, *16*, 5156–5168. [[CrossRef](#)] [[PubMed](#)]
234. Yoda, K.; Ando, S.; Okuda, A.; Kikuchi, A.; Okazaki, T. In vitro assembly of the CENP-B/ $\alpha$ -satellite DNA/core histone complex: CENP-B causes nucleosome positioning. *Genes Cells* **1998**, *3*, 533–548. [[CrossRef](#)] [[PubMed](#)]
235. Donehower, L.; Furlong, C.; Gillespie, D.; Kurnit, D. DNA sequence of baboon highly repeated DNA: Evidence for evolution by nonrandom unequal crossovers. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 2129–2133. [[CrossRef](#)] [[PubMed](#)]
236. Rubin, C.M.; Deininger, P.L.; Houck, C.M.; Schmid, C.W. A dimer satellite sequence in bonnet monkey DNA consists of distinct monomer subunits. *J. Mol. Biol.* **1980**, *136*, 151–167. [[CrossRef](#)]
237. Pike, L.M.; Carlisle, A.; Newell, C.; Hong, S.B.; Musich, P.R. Sequence and evolution of rhesus monkey alphoid DNA. *J. Mol. Evol.* **1986**, *23*, 127–137. [[CrossRef](#)]
238. Prassolov, V.S.; Kuchino, Y.; Nemoto, K.; Nishimura, S. Nucleotide sequence of the BamHI repetitive sequence, including the hindIII fundamental unit, as a possible mobile element from the Japanese monkey *Macaca fuscata*. *J. Mol. Evol.* **1986**, *23*, 200–204. [[CrossRef](#)]
239. Alves, G.; Seuánez, H.N.; Fanning, T. A Clade of New World Primates with Distinctive Alphoid Satellite DNAs. *Mol. Phylogenet. Evol.* **1998**, *9*, 220–224. [[CrossRef](#)]
240. Willard, H.F.; Wayne, J.S. Chromosome-specific subsets of human alpha satellite DNA: Analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. *J. Mol. Evol.* **1987**, *25*, 207–214. [[CrossRef](#)]
241. Alexandrov, I.A.; Mitkevich, S.P.; Yurov, Y.B. The phylogeny of human chromosome specific alpha satellites. *Chromosoma* **1988**, *96*, 443–453. [[CrossRef](#)]
242. Romanova, L.Y.; Deriagin, G.V.; Mashkova, T.D.; Tumeneva, I.G.; Mushegian, A.R.; Kisselev, L.L.; Alexandrov, I.A. Evidence for selection in evolution of alpha satellite DNA: The central role of CENP-B/p $\alpha$  binding region. *J. Mol. Biol.* **1996**, *261*, 334–340. [[CrossRef](#)]
243. Greig, G.M.; Warburton, P.E.; Willard, H.F. Organization and evolution of an alpha satellite DNA subset shared by human chromosomes 13 and 21. *J. Mol. Evol.* **1993**, *37*, 464–475. [[CrossRef](#)] [[PubMed](#)]
244. Kugou, K.; Hirai, H.; Masumoto, H.; Koga, A. Formation of functional CENP-B boxes at diverse locations in repeat units of centromeric DNA in New World monkeys. *Sci. Rep.* **2016**, *6*, 27833. [[CrossRef](#)] [[PubMed](#)]
245. Nishihara, H.; Kobayashi, N.; Kimura-Yoshida, C.; Yan, K.; Bormuth, O.; Ding, Q.; Nakanishi, A.; Sasaki, T.; Hirakawa, M.; Sumiyama, K.; et al. Coordinately Co-opted Multiple Transposable Elements Constitute an Enhancer for *wnt5a* Expression in the Mammalian Secondary Palate. *PLoS Genet.* **2016**, *12*, e1006380. [[CrossRef](#)] [[PubMed](#)]
246. Raskina, O.; Barber, J.C.; Nevo, E.; Belyayev, A. Repetitive DNA and chromosomal rearrangements: Speciation-related events in plant genomes. *Cytogenet. Genome Res.* **2008**, *120*, 351–357. [[CrossRef](#)] [[PubMed](#)]
247. Emadzade, K.; Jang, T.S.; Macas, J.; Kovařík, A.; Novák, P.; Parker, J.; Weiss-Schneeweiss, H. Differential amplification of satellite PaB6 in chromosomally hypervariable *Prospero autumnale* complex (Hyacinthaceae). *Ann. Bot.* **2014**, *114*, 1597–1608. [[CrossRef](#)] [[PubMed](#)]
248. Ershova, E.S.; Malinovskaya, E.M.; Konkova, M.S.; Veiko, R.V.; Umriukhin, P.E.; Martynov, A.V.; Kutsev, S.I.; Veiko, N.N.; Kostyuk, S.V. Copy number variation of human satellite III (1q12) with Aging. *Front. Genet.* **2019**, *10*, 704. [[CrossRef](#)]

249. Wevrick, R.; Willard, H.F. Long-range organization of tandem arrays of  $\alpha$  satellite DNA at the centromeres of human chromosomes: High-frequency array-length polymorphism and meiotic stability. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 9394–9398. [[CrossRef](#)] [[PubMed](#)]
250. Wei, K.H.C.; Grenier, J.K.; Barbash, D.A.; Clark, A.G. Correlated variation and population differentiation in satellite DNA abundance among lines of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 18793–18798. [[CrossRef](#)]
251. Kursel, L.E.; Malik, H.S. The cellular mechanisms and consequences of centromere drive. *Curr. Opin. Cell Biol.* **2018**, *52*, 58–65. [[CrossRef](#)]
252. Ruiz-Ruano, F.J.; López-León, M.D.; Cabrero, J.; Camacho, J.P.M. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* **2016**, *6*, 28333. [[CrossRef](#)]
253. Navajas-Pérez, R.; Schwarzacher, T.; De La Herrán, R.; Ruiz Rejón, C.; Ruiz Rejón, M.; Garrido-Ramos, M.A. The origin and evolution of the variability in a Y-specific satellite-DNA of *Rumex acetosa* and its relatives. *Gene* **2006**, *368*, 61–71. [[CrossRef](#)] [[PubMed](#)]
254. De La Herrán, R.; Robles, F.; Navas, J.I.; López-Flores, I.; Herrera, M.; Hachero, I.; Garrido-Ramos, M.A.; Ruiz Rejón, C.; Ruiz Rejón, M. The centromeric satellite of the wedge sole (*Dicologlossa cuneata*, Pleuronectiformes) is composed mainly of a sequence motif conserved in other vertebrate centromeric DNAs. *Cytogenet. Genome Res.* **2008**, *121*, 271–276. [[CrossRef](#)] [[PubMed](#)]
255. Fry, K.; Salser, W. Nucleotide sequences of HS- $\alpha$  satellite DNA from kangaroo rat *dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* **1977**, *12*, 1069–1084. [[CrossRef](#)]
256. Meštrović, N.; Mravinac, B.; Juan, C.; Ugarković, D.; Plohl, M. Comparative study of satellite sequences and phylogeny of five species from the genus *Palorus* (Insecta, Coleoptera). *Genome* **2000**, *43*, 776–785. [[CrossRef](#)]
257. Pons, J.; Gillespie, R.G. Evolution of satellite DNAs in a radiation of endemic Hawaiian spiders: Does concerted evolution of highly repetitive sequences reflect evolutionary history? *J. Mol. Evol.* **2004**, *59*, 632–641. [[CrossRef](#)]
258. Meštrović, N.; Mravinac, B.; Pavlek, M.; Vojvoda-Zeljko, T.; Šatović, E.; Plohl, M. Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosom. Res.* **2015**, *23*, 583–596. [[CrossRef](#)]
259. Cohen, S.; Agmon, N.; Yacobi, K.; Mislovati, M.; Segal, D. Evidence for rolling circle replication of tandem genes in *Drosophila*. *Nucleic Acids Res.* **2005**, *33*, 4519–4526. [[CrossRef](#)]
260. Cohen, S.; Agmon, N.; Sobol, O.; Segal, D. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mob. DNA* **2010**, *1*, 11. [[CrossRef](#)]
261. Šatović, E.; Vojvoda Zeljko, T.; Luchetti, A.; Mantovani, B.; Plohl, M. Adjacent sequences disclose potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network with transposable elements. *BMC Genom.* **2016**, *17*, 997. [[CrossRef](#)]
262. McGurk, M.P.; Barbash, D.A. Double insertion of transposable elements provides a substrate for the evolution of satellite DNA. *Genome Res.* **2018**, *28*, 714–725. [[CrossRef](#)]
263. Elder, J.F.; Turner, B.J. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* **1995**, *70*, 297–320. [[CrossRef](#)] [[PubMed](#)]
264. Dover, G. Molecular drive: A cohesive mode of species evolution. *Nature* **1982**, *299*, 111–117. [[CrossRef](#)] [[PubMed](#)]
265. Dover, G. Concerted evolution, molecular drive and natural selection. *Curr. Biol.* **1994**, *4*, 1165–1166. [[CrossRef](#)]
266. Feliciello, I.; Akrap, I.; Brajković, J.; Zlatar, I.; Ugarković, D. Satellite DNA as a driver of population divergence in the red flour beetle *tribolium castaneum*. *Genome Biol. Evol.* **2014**, *7*, 228–239. [[CrossRef](#)]
267. Rogers, J. In transition: Primate genomics at a time of rapid change. *ILAR J.* **2013**, *54*, 224–233. [[CrossRef](#)]
268. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szczesniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X.; et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**, *17*, 406–410. [[CrossRef](#)]
269. Jurka, J.; Walichiewicz, J.; Milosavljevic, A. Prototypic sequences for human repetitive DNA. *J. Mol. Evol.* **1992**, *35*, 286–291. [[CrossRef](#)]
270. Jurka, J.; Kapitonov, V.V.; Pavlicek, A.; Klonowski, P.; Kohany, O.; Walichiewicz, J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **2005**, *110*, 462–467. [[CrossRef](#)]
271. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, A.D. The Human Genome Browser at UCSC. *Genome Res.* **2002**, *12*, 996–1006. [[CrossRef](#)]

272. Gelfand, Y.; Rodriguez, A.; Benson, G. TRDB—The Tandem Repeats Database. *Nucleic Acids Res.* **2007**, *35*, D80–D87. [[CrossRef](#)]
273. Hubley, R.; Finn, R.D.; Clements, J.; Eddy, S.R.; Jones, T.A.; Bao, W.; Smit, A.F.A.; Wheeler, T.J. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **2016**, *44*, D81–D89. [[CrossRef](#)]
274. Ruitberg, C.M.; Reeder, D.J.; Butler, J.M. STRBase: A short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res.* **2001**, *29*, 320–322. [[CrossRef](#)]
275. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)] [[PubMed](#)]
276. Schaper, E.; Korsunsky, A.; Pečerska, J.; Messina, A.; Murri, R.; Stockinger, H.; Zoller, S.; Xenarios, I.; Anisimova, M. TRAL: Tandem repeat annotation library. *Bioinformatics* **2015**, *31*, 3051–3053. [[CrossRef](#)] [[PubMed](#)]
277. Jain, M.; Koren, S.; Miga, K.H.; Quick, J.; Rand, A.C.; Sasani, T.A.; Tyson, J.R.; Beggs, A.D.; Diltthey, A.T.; Fiddes, I.T.; et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **2018**, *36*, 338–345. [[CrossRef](#)] [[PubMed](#)]
278. Novák, P.; Neumann, P.; Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinform.* **2010**, *11*, 378. [[CrossRef](#)]
279. Novák, P.; Neumann, P.; Pech, J.; Steinhaisl, J.; MacAs, J. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **2013**, *29*, 792–793. [[CrossRef](#)]
280. Novák, P.; Robledillo, L.Á.; Koblížková, A.; Vrbová, I.; Neumann, P.; Macas, J. TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* **2017**, *45*, e111. [[CrossRef](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).