

Satellite DNA

Subjects: Biology

Contributor: Sandra Louzada, Mariana Lopes

Repetitive DNA is a major organizational component of eukaryotic genomes, being intrinsically related with their architecture and evolution. Tandemly repeated satellite DNAs (satDNAs) can be found clustered in specific heterochromatin-rich chromosomal regions, building vital structures like functional centromeres and also dispersed within euchromatin. Interestingly, despite their association to critical chromosomal structures, satDNAs are widely variable among species due to their high turnover rates. This dynamic behavior has been associated with genome plasticity and chromosome rearrangements, leading to the reshaping of genomes.

Keywords: satellite DNA ; genome architecture ; chromosome restructuring ; Robertsonian translocations ; satellite DNA transcription

1. Introduction

The linear organization of DNA sequences in the genome and how these sequences are packed into chromosomes define their architecture and influence its evolution. Repetitive DNA represents a major organizational component of eukaryotic genomes and includes sequences dispersed throughout the genome like transposable elements (TEs) and tandemly repeated sequences, such as satellite DNA (satDNA) [1][2]. Together with TEs, satDNAs contribute significantly to the differences in genome size between species, accounting for more than 50% of some species total DNA [3]. SatDNAs can be found in varied locations in the chromosomes, such as pericentromeric, subtelomeric and interstitial regions, forming blocks of constitutive heterochromatin (CH) [2][3][4][5][6][7] that are part of vital structures like centromeres and telomeres [2]. However, satDNA location is not restricted to CH with some satDNAs being found also dispersed throughout euchromatic regions in different species [5][8]. Multiple lines of evidence show that satDNAs have key roles in centromere function, heterochromatin formation and maintenance and chromosome pairing [9][10][11][12]. Interestingly, despite their association to critical chromosomal structures, satDNA families can display an astounding sequence variation even among closely related species. This results from their highly dynamic behavior, leading to rapid changes in sequence composition and array size within short evolutionary periods, which can lead to speciation (reviewed in [13]). Moreover, these sequences have been consistently correlated with fragile sites and evolutionary breakpoint regions in diverse species [14][15][16][17][18][19] and are intrinsically involved in frequent chromosomal rearrangements like Robertsonian translocations [20][21]. SatDNA dynamics has been shown to promote genome plasticity and to have an active involvement in the modulation of genomic architecture by promoting rearrangements.

Nevertheless, some satDNAs seem to have been preserved or “frozen” across different taxa during long evolutionary periods [22][23][24] with some of them being transcribed into satellite non-coding RNAs (satncRNAs). Indeed, transcripts of satDNAs have been reported in different species, highlighting a possible role for satncRNAs in the regulation of gene expression, cancer outcomes and aging [25][26][27]. This suggests that functional constraints may be causing the preservation of these sequences over the time [24][28]. Accordingly, some species centromeric satDNAs have been found to share a 17 bp motif known as the centromere protein B (CENP-B) box, representing the binding site for centromere protein B (CENP-B) [29][30]. It has been demonstrated that the CENP-B box is required for de novo centromere chromatin assembly and CENP-B protein is involved in centromere functions [31]. In this case, the conservation of a sequence motif across diverse mammalian species satDNAs [32] seems to be related to a specific function.

2. SatDNA Features and Organization in the Genome and Chromosomes: Emerging Technologies and Changing Concepts

The concept of satDNA suffered considerable changes through time. Early experiments historically coined the term “satellite DNA” referring to tandemly arranged sequences that formed satellite bands separate from the rest of the genomic DNA during density gradient centrifugation [33]. Given that no function was initially attributed to these sequences, they were considered as genomic “junk”, representing parasites proliferating independently in the genomes [34]. Today,

satDNAs are viewed as important genomic functional components. In order to understand participation of these sequences in genome architecture and evolution, we need to briefly address their organizational features, localization and mode of evolution.

SatDNA is typically organized as long arrays of head-to-tail linked repeats and usually present in the genomes in several million copies [1]. The length of the repeating unit (monomer) can range from a few base pairs up to more than 1 kb, forming arrays that may reach 100 Mb in length (reviewed by [35]), and that can form higher-order repeat (HOR) units (e.g., [36][37][38]). Human chromosome centromeres are populated by α satDNA (α SAT) organized in HORs that are structurally distinct and confer chromosome specificity [36][39]. Complex HORs have been found in non-human mammals such as insects, mouse, swine, bovids, horse, dog and elephant (reviewed in [40]), and more recently in Callitrichini monkeys [41] and Teleostei fish [42]. SatDNA arrays are mainly found clustered in heterochromatin, although studies also report the presence of short satDNA arrays dispersed along euchromatic regions [2][3][4][5][6][7]. These sequences can be found in varied locations in the chromosomes, such as pericentromeric, subtelomeric and interstitial regions [2][43][44][45], as well as being part of vital structures like centromeres and telomeres [2].

Usually more than one family of satDNAs can be found in the same genome, thus forming a library, which can be shared among closely related species. The satDNAs within the library may differ in monomer sequence, size, abundance, distribution and location (reviewed in [42]). Expansions and contractions of satDNA arrays can dramatically change the landscape of repetitive sequences, leading to significant differences of satDNA copy number among related species [46][47]. That is the case of the *Drosophila* genus, which contains very dissimilar satDNAs, varying from 0.5% in some species genomes to as high as 50% in others [48][49]. Such striking differences in satDNA abundance in *Drosophila* sp. were proposed to result predominantly from lineage-specific gains accumulated over the past 40 MY of evolution [50], ultimately causing species reproductive barriers [51][52].

The mechanisms proposed to be responsible for the amplification/deletion of repetitive DNA, consequently leading to their rapid evolutionary turnover, are unequal crossing over, replication slippage and rolling circle amplification [53]. SatDNA sequence divergence among species is quite variable, as some repeats are species-specific, while others are widely conserved, being shared across distantly related species [22][24][54]. SatDNAs have a unique mode of evolution, known as concerted evolution, a two-level process in which mutations are homogenized throughout monomers of a repetitive family and concomitantly fixed within a group of reproductively linked organisms [55][56].

The study and characterization of satDNA has lagged behind when compared with other genomic sequences. Throughout time, different methodological approaches have generated insights into the structure, organization, function and evolution of these sequence elements, although this characterization has been significantly hampered by their highly repetitive nature. The advent of high-throughput sequencing technologies and associated bioinformatics tools opened the door to whole genome sequencing projects, and as the technology became more robust and cheaper, the number of sequenced species increased exponentially. In 2018, the Earth BioGenome project was launched, aiming to increase the number of sequenced eukaryotic genomes from 2534 species (of which only 25 comply with the standard for contig and scaffold N50 established by the Genome 10K organization) to characterize the genomes of the 1.5 million known species within a 10 year time frame [57]. Of note, satDNA, as well as other repetitive sequences, have been systematically omitted from the genome projects, due to difficulties in sequence alignment and assembly, given that the read length of current sequencing technologies is unable to span the longer repeats and tandem arrays [58][59]. Nevertheless, high-throughput sequencing contributed significantly to increase our knowledge regarding satDNA sequences [60]. Next generation sequencing (NGS; e.g., Illumina), allied to newly developed bioinformatics tools capable of identifying satDNA sequences in unassembled data (e.g., RepeatExplorer) [61][62][63], helped uncover the extent of satDNAs present in the genome of different species, revealing unpredicted levels of satDNA diversity (e.g., [64][65][66][67][68][69]). For instances, 62 satDNA families were identified in the genome of the migratory locust, leading to the coining of the term 'satellitome' to refer to the whole collection of satDNA families found in a single genome [64], a part of the 'repeatome', a term proposed previously [70] to refer to the collection of all repetitive sequences in a genome (TEs, satDNAs, etc.). This number has been surpassed by a recent study where 164 satDNA families have been identified in Teleostei fish, being this the biggest satellitome characterized for a given species so far [68]. The availability of a methodology capable of assessing satDNA array abundance and diversity led to an explosion of comparative studies across a wide range of clades, including mammals, insects and plants (e.g., [41][42][67][69][71][72]) providing insights into these sequences.

The development of sequencing technologies that generate long-range data has allowed the community to overcome some of the limitations imposed by NGS and is fueling the study of repeats. Single-molecule real-time sequencing and nanopore sequencing technologies (commercialized by PacBio and Oxford Nanopore Technologies (ONT), respectively) can generate longer reads capable of spanning repetitive regions, thus enabling their assembly into contigs (reviewed in

[59]). For instances, ONT nanopore sequencers have been shown to generate unprecedented ultra-long reads that can reach mega-base lengths, leading to significant improvements in the human genome assembly [73][74][75][76], with some of the repetitive-containing gaps being closed [77][78]. By using long-read methods we are gaining access to important repeated-rich structures, like centromeres, revealing further insights into their sequence content and structure [79]. For instances, *Drosophila* centromeric satDNAs were recently shown to be intermingled with TEs [80]. Other recent studies report the improvement of human Y chromosome centromere assembly [77] and the reconstruction of a 2.8 megabase centromeric satDNA array, with the potential to achieve for the first-time telomere-to-telomere sequencing of the X chromosome [78].

Several studies demonstrate that the combination of different high-throughput sequencing methods (e.g., Illumina, ONT and PacBio) with other techniques, such as optical mapping, cytogenetics and molecular techniques, is beneficial and sometimes essential to determine satDNA features. The use of PacBio long-read sequencing together with optical mapping proved to be helpful in the assembly of satDNA arrays with large monomers and provided insights regarding recombination rates in the Eurasian crow [81]. Positional data derived from fluorescent in situ hybridization (FISH) remains vital to determine the physical location of satDNAs, since such information cannot be achieved for genomes that have not yet been properly assembled (e.g., [64][41][69][80][82]), and sequences mapping by FISH on extended DNA fibers can provide significant assistance to the process of genome assembly, aiding in contig ordering (e.g., [83][84]). Improved techniques based on FISH, helped shedding light into repetitive-rich chromosome regions with centromeric function (e.g., [85]). Other methods have also shown to provide a valid and expedite analysis of repetitive sequences profile, such as PCR-based approaches, that have been used to determine satDNA copy number differences between healthy and cancer cells/tissues [86]. In particular, the use of droplet digital PCR (ddPCR) combined with other methodologies has contributed to the validation and quantification of rare retrotransposon insertion events in different tissues including tumors [87] and the detection and accurate quantification of human *SATII* ncRNA in cancer patients [88]. The integration of genomic, cytogenetic and cell biology data helps to establish a connection between sequence information, its localization in the chromosomes and their interaction with other components of the genome, defining the field of chromosomics [89]. We believe that this approach is essential to fully understand the organization of repetitive sequences.

Other aspects of satDNA biology are also becoming accessible through the use of recent methodologies, such as the characterization of their expression and chromatin state, namely by using RNA sequencing (RNA-seq) and chromatin immunoprecipitation approaches followed by DNA sequencing (CHIP-seq) [90][91]. In particular, for CHIP-seq experiments several studies report the use of a specific antibody against DNA binding centromere-specific histone H3 (CENH3), which is an ortholog for human CENP-A. This methodology has proven to be useful for clarifying the satDNA content in the centromere, improving some organisms reference sequence and uncovering satDNA variability (e.g., [92][93]).

The data generated is now being used to determine satDNA sequences organization in the genome [94], explore predicted evolutionary patterns and hypothesis (e.g., [95][66][96][97]), as well as to shed light into the function of these sequences [80][98]. We are now closer than ever to fully access the sequence information hidden within repetitive-rich chromosome structures like centromeres and telomeres. However, we still need to further develop and adapt currently available approaches to achieve a combination of genomic, cytogenetic and molecular techniques to optimally address these regions, which we propose could be referred to as centrOMICs and telOMICs (Figure 1). SatDNAs represent one of the most intriguing and also interesting components of the genome and their full characterization will help us to better understand genome organization, architecture and evolution.

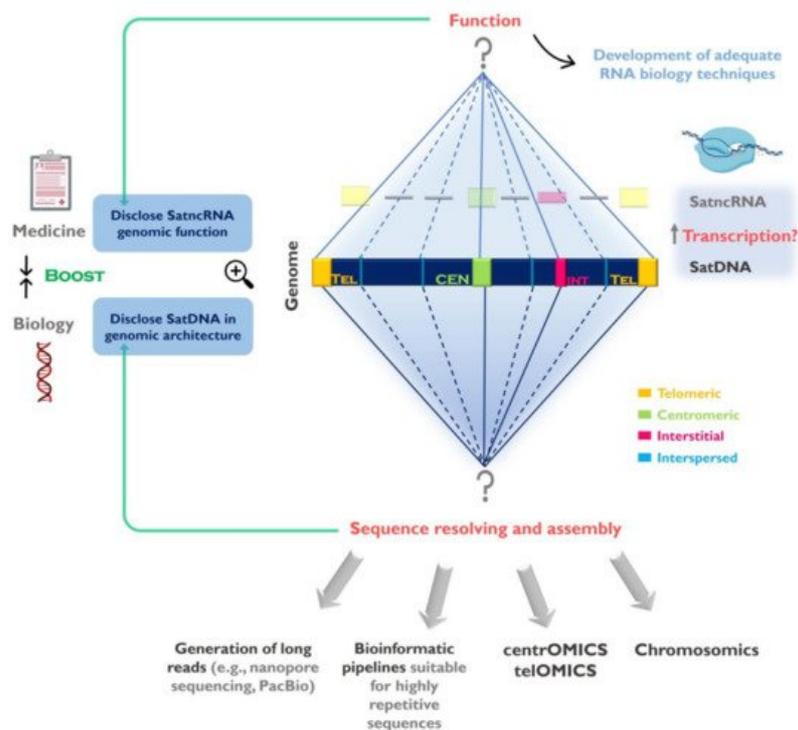


Figure 1. Challenges in the study of satellite DNA (satDNA) sequences and the importance to fully understand the repetitive genomic fraction. SatDNAs can be found clustered at the centromeres, telomeres and forming interstitial heterochromatin (CH) blocks, as well as scattered (interspersed) throughout the chromosomes. The full characterization of satDNAs needs to be addressed in two levels: 1-Disclose satDNAs linear sequence and improve their representation in genome assemblies. Despite currently used sequencing strategies (e.g., next generation sequencing (NGS)) contributed for satDNA studies, the full characterization of these sequences will only be achieved by using sequencing technologies capable of long reads, bioinformatics pipelines suitable for highly repetitive sequences, together with other techniques (e.g., FISH, optical mapping). These strategies need to be directed to specific chromosome structures such as centromeres (centrOMICS) and telomeres (telOMICS), which harbor large amounts of satDNA. Important also is the integration of genomic data with sequence localization in the chromosomes, and their interaction with other components of the genome (chromosomics); 2- Clarify satDNAs function(s) in the genome by studying the satellite non-coding RNAs (satncRNA) and their interaction with other components and structures in the genome. In this field there is the need to develop adequate biology techniques to address repetitive sequences transcription study. The disclosure of satDNA sequences will help to better understand its genomic architecture and its role in genome restructuring in evolution and disease.

References

1. Charlesworth, B.; Sniegowski, P.; Stephan, L.W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 1994, 371, 215–220.
2. Plohl, M.; Mestrovic, N.; Mravinac, B. Satellite DNA evolution. *Genome Dyn.* 2012, 7, 126–152.
3. López-Flores, I.; Garrido-Ramos, M.A. The Repetitive DNA Content of Eukaryotic Genomes. *Genome Dyn.* 2012, 7, 1–28.
4. Brajković, J.; Feliciello, I.; Bruvo-Madžarić, B.; Ugarković, D. Satellite DNA-like elements associated with genes within euchromatin of the beetle *Tribolium castaneum*. *G3 Genes Genomes Genet.* 2012, 2, 931–941.
5. Kuhn, G.C.; Küttler, H.; Moreira-Filho, O.; Heslop-Harrison, J.S. The 1.688 repetitive DNA of *Drosophila*: Concerted evolution at different genomic scales and association with genes. *Mol. Biol. Evol.* 2012, 29, 7–11.
6. Larracuenta, A.M. The organization and evolution of the Responder satellite in species of the *Drosophila melanogaster* group: Dynamic evolution of a target of meiotic drive. *BMC Evol. Biol.* 2014, 14, 233.
7. Pavlek, M.; Gelfand, Y.; Plohl, M.; Meštrović, N. Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. *DNA Res.* 2015, 22, 387–401.
8. Feliciello, I.; Akrap, I.; Ugarkovic, D. Satellite DNA modulates gene expression in the beetle *Tribolium castaneum* after heat stress. *PLoS Genet.* 2015, 11, e1005466.

9. Volpe, T.A.; Kidner, C.; Hall, I.M.; Teng, G.; Grewal, S.I.; Martienssen, R.A. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 2002, 297, 1833–1837.
10. Martienssen, R.A. Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat. Genet.* 2003, 35, 213–214.
11. Allshire, R.C.; Madhani, H.D. Ten Principles of Heterochromatin Formation and Function. *Nat. Rev. Mol. Cell Biol.* 2017, 19, 229–244.
12. Garrido-Ramos, M.A. Satellite DNA: An Evolving Topic. *Genes* 2017, 8, 230.
13. Ugarković, Đ. Evolution of α -Satellite DNA. In *Encyclopedia of Life Sciences*; John Wiley & Sons, Ltd.: Chichester, UK, 2013.
14. Adegá, F.; Guedes-Pinto, H.; Chaves, R. Satellite DNA in the karyotype evolution of domestic animals—Clinical considerations. *Cytogenet. Genome Res.* 2009, 126, 12–20.
15. Ruiz-Herrera, A.; Castresana, J.; Robinson, T.J. Is mammalian chromosomal evolution driven by regions of genome fragility? *Genome Biol.* 2006, 7.
16. Farré, M.; Bosch, M.; López-Giráldez, F.; Ponsà, M.; Ruiz-Herrera, A. Assessing the role of tandem repeats in shaping the genomic architecture of great apes. *PLoS ONE* 2011, 6, e27239.
17. Vieira-da-Silva, A.; Louzada, S.; Adegá, F.; Chaves, R. A high-resolution comparative chromosome map of *Cricetus cricetus* and *Peromyscus eremicus* reveal the involvement of constitutive heterochromatin in breakpoint regions. *Cytogenet. Genome Res.* 2015, 145, 59–67.
18. De La Fuente, R.; Baumann, C.; Viveiros, M.M. ATRX contributes to epigenetic asymmetry and silencing of major satellite transcripts in the maternal genome of the mouse embryo. *Development* 2015, 142, 1806–1817.
19. Giunta, S.; Funabiki, H. Integrity of the human centromere DNA repeats is protected by CENP-A, CENP-C, and CENP-T. *Proc. Natl. Acad. Sci. USA* 2017, 11, 1928–1933.
20. Chaves, R.; Adegá, F.; Heslop-Harrison, J.S.; Guedes-Pinto, H.; Wienberg, J. Complex satellite DNA reshuffling in the polymorphic t(1; 29) Robertsonian translocation and evolutionarily derived chromosomes in cattle. *Chromosome Res.* 2003, 11, 641–648.
21. Escudeiro, A.; Ferreira, D.; Mendes-da-Silva, A.; Heslop-Harrison, J.S.; Adegá, F.; Chaves, R. Bovine satellite DNAs—A history of the evolution of complexity and impact in the Bovidae Family. *Eur. Zool. J.* 2019, 86, 20–37.
22. Mravinac, B.; Plohl, M.; Mestrović, N.; Ugarković, D. Sequence of PRAT Satellite DNA 'Frozen' in Some Coleopteran Species. *J. Mol. Evol.* 2002, 54, 774–783.
23. Plohl, M.; Petrović, V.; Luchetti, A.; Ricci, A.; Satović, E.; Passamonti, M.; Mantovani, B. Long-term conservation vs high sequence divergence: The case of an extraordinarily old satellite DNA in bivalve mollusks. *Heredity* 2010, 104, 543–551.
24. Chaves, R.; Ferreira, D.; Mendes-da-Silva, A.; Meles, S.; Adegá, F. FA-SAT Is an Old Satellite DNA Frozen in Several Bilateria Genomes. *Genome Biol. Evol.* 2017, 9, 3073–3087.
25. Pezer, Z.; Brajković, J.; Feliciello, I.; Ugarković, D. Satellite DNA-Mediated Effects on Genome Regulation. *Genome Dyn.* 2012, 7, 153–169.
26. 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.
27. Ferreira, D.; Meles, S.; Escudeiro, A.; Mendes-da-Silva, A.; Adegá, F.; Chaves, R. Satellite non-coding RNAs: The emerging players in cells, cellular pathways and cancer. *Chromosome Res.* 2015, 23, 479–493.
28. Ferreira, D.; Escudeiro, A.; Adegá, F.; Anjo, S.I.; Manadas, B.; Chaves, R. FA-SAT ncRNA interacts with PKM2 protein: Depletion of this complex induces a switch from cell proliferation to apoptosis. *Cell Mol. Life Sci.* 2019.
29. Masumoto, H.; Masukata, H.; Muro, Y.; Nozaki, N.; Okazaki, T. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.* 1989, 109, 1963–1973.
30. Muro, Y.; Masumoto, H.; Yoda, K.; Nozaki, N.; Ohashi, M.; Okazaki, T. Centromere protein B assembles human centromeric α -satellite DNA at the 17-bp sequence, CENP-B box. *J. Cell Biol.* 1992, 116, 585–596.
31. 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.
32. Kipling, D.; Warburton, P.E. Centromeres, CENP-B and Tigger too. *Trends Genet.* 1997, 13, 141–145.
33. Kit, S.J. Equilibrium Sedimentation in Density Gradients of DNA Preparations from Animal Tissues. *J. Mol. Biol.* 1961, 3, 711–716.

34. Orgel, L.E.; Crick, F.H.C. Selfish DNA: The Ultimate Parasite. *Nature* 1980, 284, 604–607.
35. Plohl, M.; Luchetti, A.; Mestrovic, 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.
36. Willard, H.F. Chromosome-specific organization of human α satellite DNA. *Am. J. Hum. Genet.* 1985, 37, 524–532.
37. Alexandrov, I.A.; Medvedev, L.; Mashkova, T.D.; Kisselev, L.L.; Romanova, L.Y.; Yurov, Y.B. Definition of a new α satellite suprachromosomal family characterized by monomeric organization. *Nucleic Acids Res.* 1993, 21, 2209–2215.
38. McNulty, S.M.; Sullivan, B.A. α satellite DNA biology: Finding function in the recesses of the genome. *Chromosome Res.* 2018, 26, 115–138.
39. Wayne, J.S.; Willard, H.F. Chromosome-specific α satellite DNA: Nucleotide sequence analysis of the 2.0 kilobasepair repeat from the human X chromosome. *Nucleic Acids Res.* 1985, 13, 2731–2743.
40. Vlahovic, I.; Gluncic, M.; Rosandic, M.; Ugarkovic, Đ.; Paar, V. Regular Higher Order Repeat Structures in Beetle *Tribolium castaneum* Genome. *Genome Biol Evol.* 2017, 9, 2668–2680.
41. Araújo, 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.
42. Utsunomia, R.; Ruiz-Ruano, F.J.; Silva, D.M.Z.A.; Serrano, É.A.; Rosa, I.F.; Scudeler, P.E.S.; Hashimoto, D.T.; Oliveira, C.; Camacho, J.P.M.; Foresti, F. A Glimpse into the Satellite DNA Library in Characidae Fish (Teleostei, Characiformes). *Front. Genet.* 2017, 8, 103.
43. Adegá, F.; Chaves, R.; Guedes-Pintos, H. Chromosomal evolution and phylogenetic analysis in *Tayassu pecari* and *Pecari tajacu* (Tayassuidae): Tales from constitutive heterochromatin. *J. Genet.* 2007, 86, 19–26.
44. Henikoff, S.; Dalal, Y. Centromeric chromatin: What makes it unique? *Curr. Opin. Genet. Dev.* 2005, 15, 177–184.
45. Paço, A.; Adegá, F.; Guedes-Pinto, H.; Chaves, R. Hidden heterochromatin Characterization in the Rodentia species *Cricetus cricetus*, *Peromyscus eremicus* (Cricetidae) and *Praomys tullbergi* (Muridae). *Genet. Mol. Biol.* 2009, 32, 58–68.
46. Ugarković, Đ.; Plohl, M. Variation in satellite DNA profiles—Causes and effects. *EMBO J.* 2002, 21, 5955–5959.
47. Louzada, S.; Vieira-da-Silva, A.; Mendes-da-Silva, A.; Kubickova, S.; Rubes, J.; Adegá, F.; Chaves, R. A Novel Satellite DNA Sequence in the *Peromyscus* Genome (PMSat): Evolution via Copy Number Fluctuation. *Mol. Phylogenet. Evol.* 2015, 92, 193–203.
48. Gall, J.G.; Cohen, E.G.; Polan, M.L. Repetitive DNA sequences in *Drosophila*. *Chromosoma* 1971, 33, 319–344.
49. Lohe, A.R.; Brutlag, D.L. Identical satellite DNA sequences in sibling species of *Drosophila*. *J. Mol. Biol.* 1987, 194, 161–170.
50. Wei, K.H.-C.; Lower, S.E.; Caldas, I.V.; Sless, T.J.S.; Barbash, D.A.; Clark, A.G. Variable rates of simple satellite gains across the *Drosophila* phylogeny. *Mol. Biol. Evol.* 2018, 35, 925–941.
51. Ferree, P.M.; Barbash, D.A. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol.* 2009, 7, e1000234.
52. Ferree, P.M.; Prasad, S. How can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. *Genet Res Int* 2012, 430136.
53. Walsh, J.B. Persistence of tandem arrays: Implications for satellite and simple-sequence DNAs. *Genetics* 1987, 115, 553–567.
54. Adegá, F.; Chaves, R.; Guedes-Pinto, H. Suiformes orthologous satellite DNAs as a hallmark of collared and white-lipped peccaries (Tayassuidae) evolutionary rearrangements. *Micron* 2008, 39, 1281–1287.
55. Dover, G.A. Molecular drive in multigene families: How biological novelties arise, spread and are assimilated. *Trends Genet.* 1986, 2, 159–165.
56. Elder, J.F., Jr.; Turner, B.J. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* 1995, 70, 297–320.
57. Lewin, H.A.; Robinson, G.E.; Kress, W.J.; Baker, W.J.; Coddington, J.; Crandall, K.A.; Durbin, R.; Edwards, S.V.; Forest, F.; Gilbert, M.T.P.; et al. Earth BioGenome Project: Sequencing life for the future of life. *Proc. Natl. Acad. Sci. USA* 2018, 115, 4325–4333.
58. Treangen, T.J.; Salzberg, S.L. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nat. Rev. Genet.* 2011, 13, 36–46.

59. Van Dijk, E.L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. The Third Revolution in Sequencing Technology. *Trends Genet.* 2018, 34, 666–681.
60. Sedlazeck, F.J.; Lee, H.; Darby, C.A.; Schartz, M.C. Piercing the dark matter: Bioinformatics of long-range sequencing and mapping. *Nat. Rev. Genet.* 2018, 19, 329–346.
61. 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.
62. 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.
63. Salmela, L.; Rivals, E. LoRDEC: Accurate and efficient long read error correction. *Bioinformatics* 2014, 30, 3506–3514.
64. 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.
65. Cook, D.E.; Zdraljevic, S.; Tanny, R.E.; Seo, B.; Riccardi, D.D.; Noble, L.M.; Rockman, M.V.; Alkema, M.J.; Braendle, C.; Kammenga, J.E.; et al. The Genetic Basis of Natural Variation in *Caenorhabditis elegans* Telomere Length. *Genetics* 2016, 204, 371–383.
66. Wei, K.H.; 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.
67. Silva, D.M.Z.A.; Utsunomia, R.; Ruiz-Ruano, F.J.; Daniel, S.N.; Porto-Foresti, F.; Hashimoto, D.T.; Oliveira, C.; Camacho, J.P.M.; Foresti, F. High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Sci. Rep.* 2017, 7, 12726.
68. Utsunomia, R.; Silva, D.M.Z.A.; Ruiz-Ruano, F.J.; Goes, C.A.G.; Melo, S.; Ramos, L.P.; Oliveira, C.; Porto-Foresti, F.; Foresti, F.; Hashimoto, D.T. Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Sci. Rep.* 2019, 9, 5856.
69. Palacios-Gimenez, O.M.; Dias, G.B.; De Lima, L.G.; Ramos, É.; Martins, C.; Cabral-de-Mello, D.C. High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket *Eneoptera surinamensis*. *Sci. Rep.* 2017, 7, 6422.
70. Kim, Y.B.; Oh, J.H.; McIver, L.J.; Rashkovetsky, E.; Michalak, K.; Garner, H.R.; Kang, L.; Nevo, E.; Korol, A.B.; Michalak, P. Divergence of *Drosophila melanogaster* repeatomes in response to a sharp microclimate contrast in Evolution Canyon, Israel. *Proc. Natl. Acad. Sci. USA* 2014, 111, 10630–10635.
71. Usai, G.; Mascagni, F.; Natali, L.; Giordani, T.; Cavallini, A. Comparative Genome-Wide Analysis of Repetitive DNA in the Genus *Populus* L. *Tree Genet. Genomes* 2017, 13, 96.
72. Vozdova, M.; Kubickova, S.; Cernohorska, H.; Fröhlich, J.; Rubes, J. Satellite DNA Sequences in Canidae and Their Chromosome Distribution in Dog and Red Fox. *Cytogenet. Genome Res.* 2017, 150, 118–127.
73. Miga, K.H. Completing the human genome: The progress and challenge of satellite DNA assembly. *Chromosome Res.* 2015, 23, 421–426.
74. Jain, M.; Olsen, H.E.; Paten, B.; Akeson, M. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol.* 2016, 17, 239.
75. Kuderna, L.F.K.; Lizano, E.; Julià, E.; Gomez-Garrido, J.; Serres-Armero, A.; Kuhlwilm, M.; Alandes, R.A.; Alvarez-Estape, M.; Juan, D.; Simon, H.; et al. Selective Single Molecule Sequencing and Assembly of a Human Y Chromosome of African Origin. *Nat. Commun.* 2019, 10, 4.
76. Miga, K.H. Centromeric Satellite DNAs: Hidden Sequence Variation in the Human Population. *Genes* 2019, 10, 352.
77. Jain, M.; Koren, S.; Miga, K.H.; Quick, J.; Rand, A.C.; Sasani, T.A.; Tyson, J.R.; Beggs, A.D.; Dilthey, 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.
78. 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. *bioRxiv* 2019, 735928.
79. Muller, H.; Gil, J., Jr.; Drinnenberg, I.A. The impact of centromeres on spatial genome architecture. *Trends Genet.* 2019, 35, 565–578.
80. Chang, C.H.; Chavan, A.; Palladino, J.; Wei, X.; Martins, N.M.C.; Santinello, B.; Chen, C.C.; Erceg, J.; Beliveau, B.J.; Wu, C.T.; et al. Islands of retroelements are major components of *Drosophila* centromeres. *PLoS Biol.* 2019, 17, e3000241.

81. Weissensteiner, M.H.; Pang, A.W.C.; Bunikis, I.; Höijer, I.; Vinnere-Petterson, O.; Suh, A.; Wolf, J.B.W. Combination of Short-Read, Long-Read, and Optical Mapping Assemblies Reveals Large-Scale Tandem Repeat Arrays with Population Genetic Implications. *Genome Res.* 2017, 27, 697–708.
82. Ruiz-Ruano, F.J.; Cuadrado, Á.; Montiel, E.E.; Camacho, J.P.; López-León, M.D. Next Generation Sequencing and FISH Reveal Uneven and Nonrandom Microsatellite Distribution in Two Grasshopper Genomes. *Chromosoma* 2015, 124, 221–234.
83. Skinner, B.M.; Sargent, C.A.; Churcher, C.; Hunt, T.; Herrero, J.; Loveland, J.E.; Dunn, M.; Louzada, S.; Fu, B.; Chow, W.; et al. The Pig X and Y Chromosomes: Structure, Sequence, and Evolution. *Genome Res.* 2016, 26, 130–139.
84. Louzada, S.; Komatsu, J.; Yang, F. Fluorescence in situ hybridization onto DNA fibres generated using molecular combing. In *Fluorescence In Situ Hybridization (FISH)*; Liehr, T., Ed.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 275–293.
85. Giunta, S. Centromere Chromosome Orientation Fluorescent In Situ Hybridization (Cen-CO-FISH) Detects Sister Chromatid Exchange at the Centromere in Human Cells. *Bio-protocol* 2018, 8.
86. Saha, A.; Mourad, M.; Kaplan, M.H.; Chefetz, I.; Malek, S.N.; Buckanovich, R.; Markovitz, D.M.; Contreras-Galindo, R. The Genomic Landscape of Centromeres in Cancers. *bioRxiv* 2019, 505800.
87. White, T.B.; McCoy, A.M.; Streva, V.A.; Fenrich, J.; Deininger, P.L. A droplet digital PCR detection method for rare L1 insertions in tumors. *Mobile DNA* 2014, 5, 30.
88. Kishikawa, T.; Otsuka, M.; Yoshikawa, T.; Ohno, M.; Yamamoto, K.; Yamamoto, N.; Yamamoto, N.; Kotani, A.; Koike, K. Quantitation of circulating satellite RNAs in pancreatic cancer patients. *JCI Insight* 2016, 1, e86646.
89. Deakin, J.E.; Potter, S.; O'Neill, R.; Ruiz-Herrera, A.; Cioffi, M.B.; Eldridge, M.D.B.; Fukui, K.; Marshall Graves, J.A.; Griffin, D.; Grutzner, F.; et al. Chromosomics: Bridging the Gap between Genomes and Chromosomes. *Genes* 2019, 10, 627.
90. Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.Y.; Schones, D.E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* 2007, 129, 823–837.
91. 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.
92. Kowar, T.; Zakrzewski, F.; Macas, J.; Koblížková, A.; Viehovec, P.; Weisshaar, B.; Schmidt, T. Repeat Composition of CenH3-chromatin and H3K9me2-marked heterochromatin in Sugar Beet (*β vulgaris*). *BMC Plant Biol.* 2016, 16, 120.
93. Su, H.; Liu, Y.; Liu, C.; Shi, Q.; Huang, Y.; Han, F. Centromere satellite repeats have undergone rapid changes in polyploid wheat subgenomes. *Plant Cell* 2019, 31, 2035–2051.
94. Khost, D.E.; Eickbush, D.G.; Larracuenta, A.M. Single-Molecule Sequencing Resolves the Detailed Structure of Complex Satellite DNA Loci in *Drosophila*. *Genome Res.* 2017, 27, 709–721.
95. Belyayev, A.; Josefiová, J.; Jandová, M.; Kalendar, R.; Krak, K.; Mandák, B. Natural History of a Satellite DNA Family: From the Ancestral Genome Component to Species-Specific Sequences, Concerted and Non-Concerted Evolution. *Int. J. Mol. Sci.* 2019, 20, 1201.
96. De Lima, L.G.; Svartman, M.; Kuhn, C.S. Dissecting the Satellite DNA Landscape in Three Cactophilic *Drosophila* Sequenced Genomes. *G3 Genes Genomes Genet.* 2017, 7, 2831–2843.
97. Flynn, J.M.; Long, M.; Wing, R.A.; Clark, A.G. Evolutionary dynamics of abundant 7 bp satellites in the genome of *Drosophila virilis*. *bioRxiv* 2019, 693077.
98. Hartley, G.; O'Neill, R.J. Centromere Repeats: Hidden Gems of the Genome. *Genes* 2019, 10, 223.