



Review

# Conversion of DNA Sequences: From a Transposable Element to a Tandem Repeat or to a Gene

Ana Paço <sup>1,\*</sup> , Renata Freitas <sup>2,3,4</sup>  and Ana Vieira-da-Silva <sup>1</sup>

<sup>1</sup> MED-Mediterranean Institute for Agriculture, Environment and Development, University of Évora, 7002–554 Évora, Portugal; ana.vieiradsilva@gmail.com

<sup>2</sup> IBMC-Institute for Molecular and Cell Biology, University of Porto, R. Campo Alegre 823, 4150–180 Porto, Portugal; Renata.Freitas@ibmc.up.pt

<sup>3</sup> I3S-Institute for Innovation and Health Research, University of Porto, Rua Alfredo Allen, 208, 4200–135 Porto, Portugal

<sup>4</sup> ICBAS-Institute of Biomedical Sciences Abel Salazar, University of Porto, 4050-313 Porto, Portugal

\* Correspondence: apaco@uevora.pt; Tel.: +351-266-760-878

Received: 19 October 2019; Accepted: 29 November 2019; Published: 5 December 2019



**Abstract:** Eukaryotic genomes are rich in repetitive DNA sequences grouped in two classes regarding their genomic organization: tandem repeats and dispersed repeats. In tandem repeats, copies of a short DNA sequence are positioned one after another within the genome, while in dispersed repeats, these copies are randomly distributed. In this review we provide evidence that both tandem and dispersed repeats can have a similar organization, which leads us to suggest an update to their classification based on the sequence features, concretely regarding the presence or absence of retrotransposons/transposon specific domains. In addition, we analyze several studies that show that a repetitive element can be remodeled into repetitive non-coding or coding sequences, suggesting (1) an evolutionary relationship among DNA sequences, and (2) that the evolution of the genomes involved frequent repetitive sequence reshuffling, a process that we have designated as a “DNA remodeling mechanism”. The alternative classification of the repetitive DNA sequences here proposed will provide a novel theoretical framework that recognizes the importance of DNA remodeling for the evolution and plasticity of eukaryotic genomes.

**Keywords:** tandem repeats; dispersed sequences; origin of tandem repeats; mobilization of tandem repeats; DNA remodeling mechanism

## 1. Introduction

Eukaryotic genomes contain a high diversity of repetitive DNA sequences [1,2]. The amplification/deletion of these sequences contributed significantly to the extraordinary variation in genome size found between taxa [3–7]). In the animal kingdom, the genome size could vary from 20 Mb to 130 Gb, which is mainly due to differences in the content of repetitive sequences [8]. The same large variation was identified in plants. For instance, the fraction of repetitive genomic DNA is 13–14% (125 Mb–157 Mb) in *Arabidopsis thaliana* but, contrastingly, is 77% (2.5 GB) in *Zea mays* [9].

The biological role of this repetitive DNA fraction has been a topic of great interest namely to understand evolution and disease [10–13]. In summary, this fraction seems to be involved in DNA packaging, the evolutionary events of the genome (through promoting DNA instability), gene expression, and epigenetic mechanisms [14–25]. The analysis of the DNA sequences located in chromosomal breakpoint regions strongly suggests that repetitive sequences work as a driving force in the occurrence of chromosomal rearrangements, since these regions are extremely rich in these sequences [26,27]. The repetitive nature, per se, of the different classes of repeats located in these

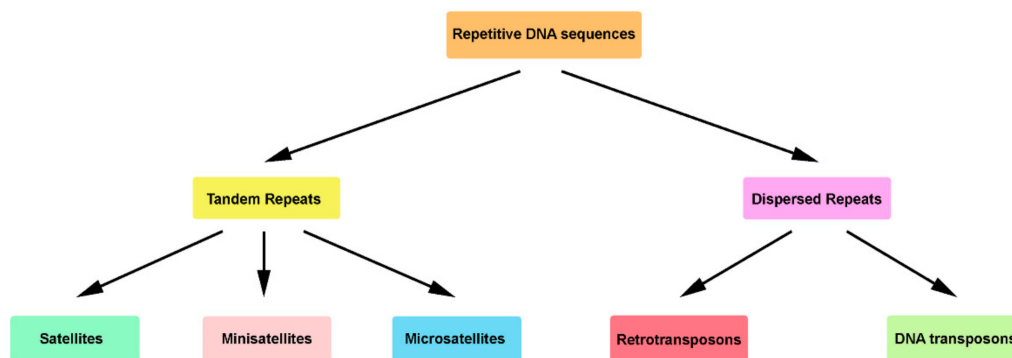
breakpoint regions favours recombinational events between homologous sequences in non-homologous regions, which may culminate in chromosomal restructurings [21]. Besides, an analysis devoted to the transcriptional activity of some repetitive sequences input a role of these sequences in control of gene expression, cellular response to stress and centromeric function, specifically by RNA interference mechanisms [28].

Despite the increasing evidence pointing to a functional significance of the repetitive DNA fraction, there are still limitations to characterizing their role in different biological processes due to their high diversity, genomic abundance, complex evolution mode, and difficulty in isolation and sequencing [29–31]. Thus, the real biological relevance of this fraction of eukaryotic genomes is yet to be revealed.

This review presents a compilation of evidence on the organization and evolutionary relationship between different classes of repetitive sequences, and between repetitive sequences and genes. This evidence shows that the classification of repetitive sequences based on its genomic organization as “in tandem” or “dispersed repeats” is not “as black and white”, which reinforces the need for an updated classification. Further, this evidence also shows a high remodulation of the repetitive sequences in the eukaryotic genomes, which could culminate in the origin of a new sequence type, namely genes. This new way to face the classification and study of the repetitive sequences will contribute to a better understanding of genomic plasticity and its contribution to eukaryotic species evolution and adaptation to environment.

## 2. Tandem Repeats and Dispersed Repeats: Is Their Organization So Different?

Repetitive sequences are classified in two classes, tandem and dispersed repeats, according to the genomic organization of their copies (Figure 1) [32,33]. In turn, each class is divided into several subclasses or families [3–6,34–37]. With the goal to facilitate the annotation and classification of the repetitive elements in eukaryotic genomes, an open access database for repetitive sequence families (Dfam) was built [37,38].



**Figure 1.** Repetitive DNA sequences in eukaryotic genomes. This schematization collects the information of several works [16,20,33,39–42]. Here, only the largest subclasses of tandem and dispersed repeats are represented, not including the genic repetitive DNA sequences families, as tandem paralogues genes, ribosomal genes (tandem organization), retropseudogenes, transfer RNA genes, and dispersed paralogues genes (dispersed organization).

### 2.1. Tandem Repeats Organization

Traditionally, tandem repeats have been structurally characterized by a sequential arrangement of repeat units, positioned one after the other in two possible repeat orientations, head-to-tail repeats (direct repeats) or head-to-head repeats (inverted repeats) [33]. Excluding the genic repetitive DNA sequences (e.g., ribosomal genes), three distinct subclasses are mainly recognized, namely microsatellites, minisatellites, and satellites (satellite DNAs—satDNAs; nongenic tandem repeats) (Figure 1).

It is generally considered that the main difference between micro, mini, and satellites relates to the length of their array of repeat units in a chromosomal location [33,39,43]. Micro and minisatellites are classified as short tandem repeats, composed by shorter arrays of copies: ranging from 10 to 100 repeat units for microsatellites, and up to 100 repeat units for minisatellites [39,44]. However, it is important to mention that there is no consensus on the definition of microsatellites and minisatellites, since there is also no consensus in the repeat unit size, or in the minimal number of repeat units in an array of copies differentiating micro and minisatellites. The threshold for repeat unit size varies between six and 10 nucleotides [33,39,45].

The SatDNAs have traditionally been considered as organized into long arrays, with millions of copies and, thus, have been named as long tandem repeats [39,46]. However, satDNAs with a different organization have already been reported. Louzada and colleagues (2015) [47], identify a satDNA (PMSat) with high sequence conservation in the genomes of five rodents. However, the PMSat is not always organized into long array repeat units, even in species where it is highly abundant, as is the case of *Peromyscus maniculatus bairdii*. In this species the authors found short PMSat arrays and even dispersed isolated monomers, which reinforces that the limitation of techniques to isolate and analyse the repetitive fractions of genomes (as sequencing, assembly and mapping technologies) is a determining factor for the study of repetitive sequences, as well as for its accurate classification. Other reports have also come to prove the same, showing a dispersed genomic organization of short arrays of satDNA repeat units [44,48].

Besides the length of arrays, the repeat units of satDNA (monomers) can also show a great variation in size, ranking from five nucleotides, as in human satellite III [49], which is similar to a micro or a minisatellite, but with longer arrays of repeat units, up to several hundred base pairs as in the *Microtus* MSAT-2570 [50]. However, for plants and animals, the most common length is 150–180 bp and 300–360 bp, respectively, which is believed to be associated with the requirements of DNA length wrapped around one or two nucleosomes [11,51,52].

The genomic distribution presented by micro, mini, and satDNAs is also traditionally considered distinct in the literature. Generally, both microsatellites and minisatellites are distributed throughout the genome (dispersed), in both euchromatic and heterochromatic regions [33,45]. Nevertheless, minisatellites are also characterized by their accumulations in (sub)telomeric regions [43,53]. The satDNAs are mainly located in heterochromatic regions of the chromosomes, thus satDNAs are preferentially found in and around centromeres [54–57]. However, once again, exceptions have been reported as to what is generally assumed. Indeed, satDNA can also be located at interstitial and terminal positions of chromosomes [26,47,48].

## 2.2. Dispersed Repeats Organization

Dispersed repeats are mainly represented by transposable elements (TEs). With the recent proliferation of genomic sequencing studies, TEs have emerged as highly diverse, ubiquitous and abundant genomic elements, constituting approximately half of the human genome and up to 95% of DNA in plants [58–61]. The diversity of TEs reflects their evolutionary mode [62]. By the accumulation of mutations, TEs generate new families and subfamilies, “escaping” to selection [63]. Despite this diversity, it was possible to group them into two major classes according to their modes of transposition (mobilization in genomes): retrotransposons (RE, class I elements) and DNA transposons (class II elements) (Figure 1) [60,61].

Traditionally, the dispersed repeats consist of sequences represented several times in the genome, whose copies are not clustered or are organized in short clusters, presenting a wide distribution throughout the genome [39,43,64]. However, some works show dispersed repeats with an accumulated genomic organization [65–70], which can lead to the supposition of some tandem organization for these sequences. As referred to previously, the genomic accumulation of tandem repeats is common in certain genomic locations, such as the telomeric sequences (microsatellites of (TTAGGG/CCCTAA)<sub>n</sub>) in the genomes of all mammals at telomeric and interstitial positions [71–73], and satDNAs at centromeric

regions [57]. In fact, different works have specifically explored the organization of tandem repeats at telomeric and centromeric regions [26,66]. Nevertheless, how TE blocks accumulated in certain regions of the genome are in fact organized is not reported in many studies. Only a few works have revealed that TE clusters could be interrupted by other TEs or also by genes, where there might also be a short tandem arrangement of some TEs [67,69].

### 2.3. Repetitive Sequences: A New Classification Based on the Presence or Absence of Retrotransposons/Transposon Specific Domains

In this review, we suggest an alternative to the traditional classification of DNA repetitive sequences mainly based on the genomic organization of its copies. We suggest that the repetitive sequences should not be divided as tandem repeats and dispersed repeats, once repeats organized in tandem could also present a dispersed organization, and vice-versa. As referred to previously, several tandem repeats have a dispersed organization in distinct genomes. In contraposition, some TE copies may be positioned one after the other in tandem organization, presenting some chromosomal regions several complete or incomplete copies of a specific TE. As such, these sequences should be classified regarding other characteristics, namely the presence or absence of retrotransposons/transposon specific domains in the sequence of its copies. Therefore, the repetitive sequences could perhaps be classified as repeats presenting retrotransposons/transposon specific domains or repeats not presenting retrotransposons/transposon specific domains.

## 3. Repetitive DNA Remodelling

Different studies show that a repetitive element can be remodelled into a different sequence, repetitive or not, which proves an evolutionary relationship among DNA sequences, and suggests that the evolution of the genome is characterized by a frequent repetitive sequence reshuffling, a process that we have called a “DNA remodelling mechanism”. In fact, some authors show that the tandem repeats could have their origin in TEs [74], and that satDNAs could also evolve to coding sequences [75].

### 3.1. Transposable Elements in Origin and Genomic Distribution of Micro, Mini and Satellite DNAs

Sequence similarities between tandem repeats and TEs [76–78] indicate a strong evolutionary relationship between these repetitive sequences. In addition, it is also believed that TEs are involved in the origin of some sequence motifs that characterized some satDNAs, as the CENP-B box, presenting this sequence motif strong similarity with the terminal inverted repeats of pogo transposons [79]. In fact, computer simulations have suggested that satDNA monomers could be generated from a wide variety of non-satellite sequences and propagated into an array by unequal crossing-over [80]. These non-satellite sequences are often TEs [26,31,76,81–88]. In Table 1, different examples known in the literature are listed where TEs or parts of TEs were converted into other repetitive sequence or altered genes.

**Table 1.** Transposable elements in the origin of other repetitive sequence or altered genes.

Transposable Element	New Sequence or New Sequence Variable	Reference
SINE-like elements	Satellite 1 of <i>Xenopus leavis</i>	[89]
LTR retrotransposons	RPCS satDNA of <i>Ctenomys</i> rodents	[81]
SINE-like elements	Hy/Pol III satDNA european salamander	[82]
pDv mobile element	pvB370 satDNA of <i>Drosophyla virilis</i>	[83]
LINE-1 elements	Common cetacean satDNA	[76]
TART and HeT-A retrotransposons	18HT satDNA of <i>Drosophila melanogaster</i>	[90]
Atenspm2 transposons	Ensat1 of <i>Arabidopsis thaliana</i>	[84]
Crwydryn retrotransposon	E3900 satDNA of rye	[91]
MITE elements	D1100 satDNA of rye	[91]
SGM-IS transposons	SGM satDNA <i>Drosophila guanche</i>	[92]

Table 1. Cont.

Transposable Element	New Sequence or New Sequence Variable	Reference
Ty3/gypsy-retroelement	250 elements of satDNA of wheat	[93]
MITE-like elements	Xstir satDNA of <i>Xenopus leavis</i>	[94]
MITE elements	<i>HindIII</i> satDNA of oysters	[95]
Sore1 retrotransposon	Sobo satDNA of potatoes	[96]
CR1 retrotransposons	<i>HinfI</i> satDNA of chicken	[97]
Ty3/gypsy-like ogre elements	PisTR-A satDNA of pea	[31]
MITE-like elements	BIV160 satDNA of bivalves	[77]
CR1-C retrotransposons	Cen2, 3, 4, 7 and Cen11 satDNAs of chicken	[98]
CRM1 and CRM4 retrotransposons	CRM1TR satDNA of maize	[87]
Helytrons elements	CTRs satDNA of <i>Drosophyla</i>	[88]
LINE-1 elements	PROsat of <i>Phodopus roborovskii</i>	[26]
<i>Alu</i> elements	A-rich primates' microsatellites	[99]
SINE elements	BARE-1, WIS2-1A and PREM1 microsatellites of barley	[100]
LINE-1 elements	A-rich mammalian microsatellites	[101]
<i>Alu</i> elements	(GAA) <sub>n</sub> human microsatellite	[102]
MITE elements	GTCY(n) microsatellites of insects	[86]
<i>Alu</i> elements	pλg3 human minisatellite	[103]
MaLR retrotransposon	Ms6-hm mouse minisatellites	[104]
SINE B1 elements	(GGCAGA) <sub>n</sub> mouse minisatellite	[105]
<i>Alu</i> elements	(CGGGAGGC) <sub>n</sub> human minisatellite	[106]
<i>Alu</i> elements	Minisatellites of human	[107]
Gmr9/Gm ogre retrotransposons	Gmr9-associated minisatellites of soybean	[108]
LINE-1 elements	TRIM5 gene with a cyclophilin A domain	[109]

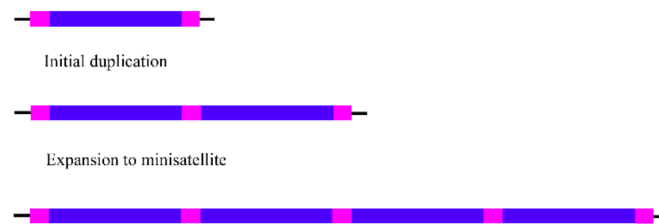
SINE: Short interspersed nuclear element, LTR: Long terminal repeats retrotransposons, LINE: Long interspersed nuclear element, MITE: Miniature inverted repeat transposable elements, satDNA: Satellite DNA.

The evidence for TE conversion into new non-coding repetitive sequences or genes are reinforced by the fact that, in humans and mice—the first fully-sequenced genomes—it was estimated that the repetitive DNA derived from TEs comprises from 40% to almost half of these genomes [110,111]. These values could be quite underestimated, with substantial amounts of older sequences not being detected due to their already highest divergence compared to the consensus sequences used for their detection. Ahmed and Liang (2012) [107], for example, considered that the ability of TEs to contribute to genome expansion is due, not only to retrotransposition (increasing its copy number in the genomes), but also by generating tandem repeats.

The exact mechanisms underlying the origin/expansion of tandem repeats from TEs are not yet completely known, but probably involved more than one mechanism. A tandem repeat sequence arises after amplification events followed by subsequent molecular mechanisms. It is widely accepted that the first repetitions of microsatellites may have originated by chance, and then expanded by slipped-strand mispairing, as proposed by Levinson and Gutman (1987) [112]. However, some studies suggested that TEs contain one or more sites predisposed to the formation of microsatellites. The Poly(A) tract at the 3' end of mammalian non-LTR retrotransposons (autonomous LINES/nonautonomous SINEs) provides a susceptible site to reverse transcription errors, which could lead to the genesis of A-rich microsatellites [113]. The description of microsatellites located at the 5' end and internal regions of retroelements is also available in the literature [100].

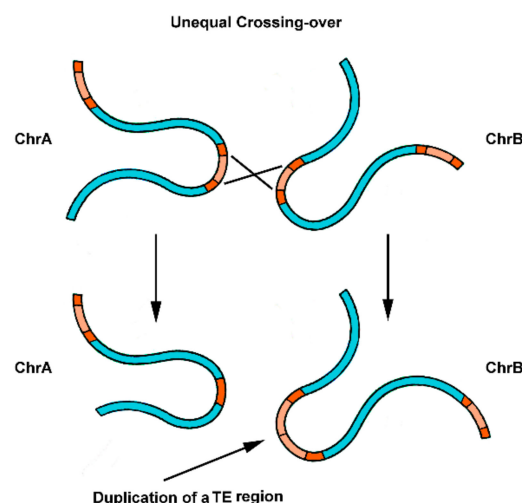
Regarding minisatellites, several reports point to their origin from a variety of TE families and subfamilies [103–106,108], namely from nonautonomous non-LTR retrotransposons as *Alu* and B1 SINE elements [105,106] or from LTR retrotransposons [103,104,108]. According to Haber and Louis (1998) [114], the origin (initial event) of the first repetitions of some minisatellites appears to have been mediated by replication slippage or unequal crossing-over, involving very short repeats (5–10 bp) that flank a motif which will be amplified as the repetition unit of these minisatellites (Figure 2). The evidence that repetitive elements, such as *Alu* elements, commonly have short direct repeats in their sequences, makes them very prone to the origin of minisatellites by this mechanism [106].

Subsequently, the amplification of the duplicated motifs into a minisatellite array could then occur by additional replication slippage events, gene conversion, or by unequal crossing-over between the longer homologous regions [114]. This mechanism seems plausible to explain the origin of minisatellites; however, it cannot explain the origin of the satDNA with larger repeat units, due to the distance over which the initial event must have occurred (replication slippage or unequal crossing-over involving flanking short repeats). Nevertheless, a very similar mechanism was proposed to explain the origin of satDNAs, as for maize centromeric repeats, with most of their monomers presenting more than 700 bp [87].



**Figure 2.** Initial event in the origin of the first minisatellites repetitions. Origin of a duplication by replication slippage or unequal crossing-over between short flanking repeats, followed by a subsequent expansion into a minisatellite.

Other interesting theories on the origin of satDNAs from TEs have been suggested. Wong and Choo (2004) [74] proposed the “first steps” hypothesis for the origin of satDNA repetitions, based on the duplication of part of a TE sequence by unequal crossing-over between homologous TE elements, which could be in the same or in different chromosomes (Figure 3). Once a tandem repetition of full or partial TEs is generated in a genome, the expansion of these novel repeat units can slowly occur over time. Mutational changes, followed by successive rounds of crossing-over homogenization (concerted evolution of tandem repeats), can justify the divergence observed between the emergent satDNA and the original TE, presenting only conserved parts of their sequences [74]. This mechanism is recurrently used to explain the origin of satDNAs. An example is the work of Dias et al. (2015) [88], suggesting the emergence of a satDNA from central tandem repeats of a helitron (DINE-TR1) in *Drosophila* species.

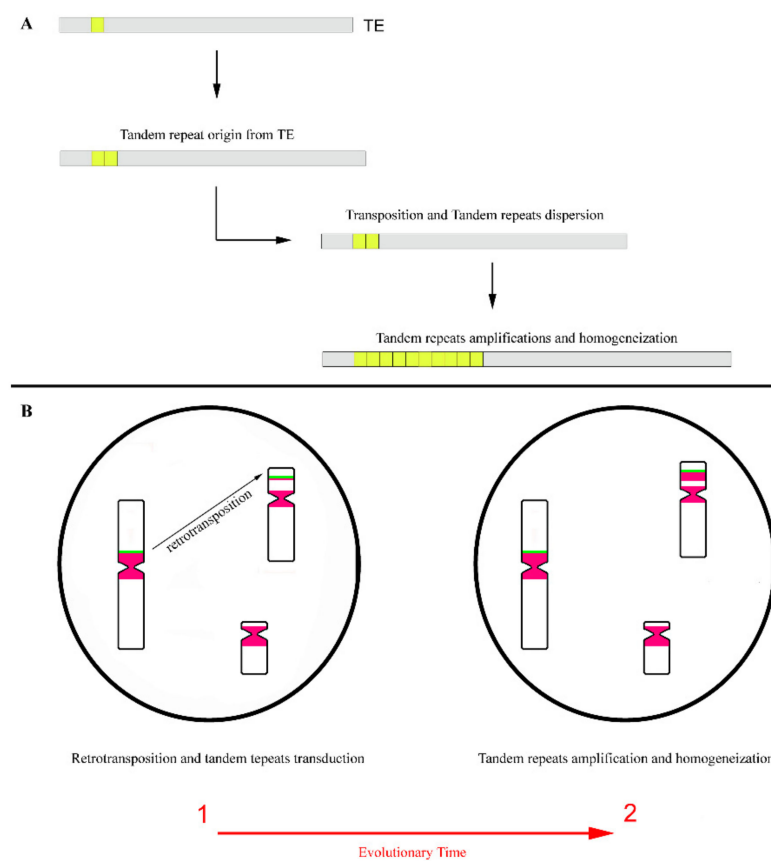


**Figure 3.** Initial steps for the origin of satDNA repeats from parts of a TE. The duplication of part of a TE sequence occurs by unequal crossing-over between homologous dispersed repeats present in chromosomes A and B (chrA and ChrB). The expansion of these novel repeat units can occur through time and result in a satDNA array of copies.

The DNA transposons, or specifically their transposase activity, have been also referred to in the birth of sequence duplications. Kapitonov and Jurka (1999) [84] propose that the breaks induced by

transposases during transposition (endonucleolytic transposase activity) could favour recombination processes in order to repair the double strand breaks. This event may originate the first repetitions of a tandem repeat, which could afterwards be amplified in a large array of copies.

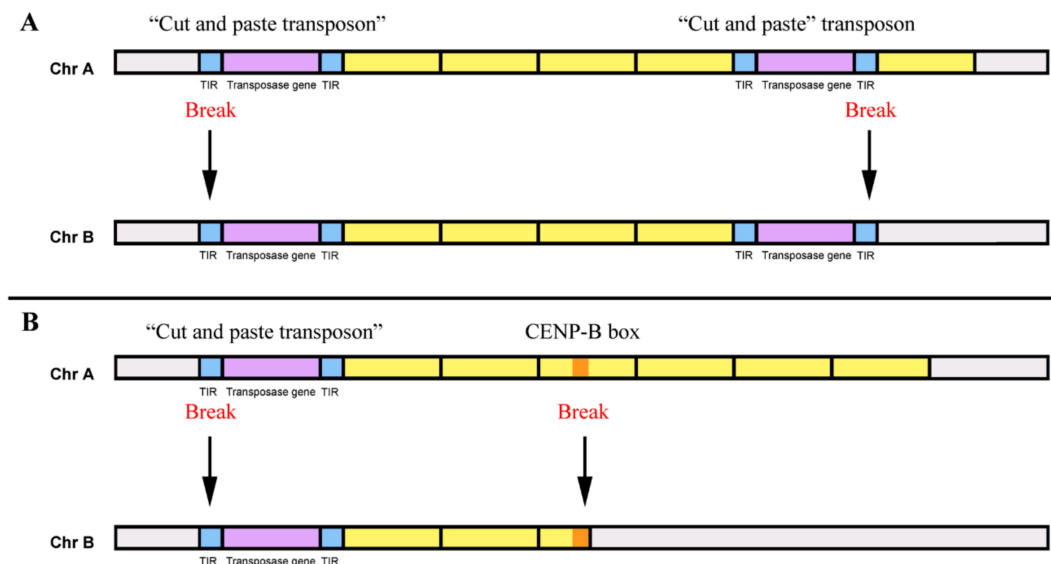
Beyond the role suggested in the origin of the tandem repeat, the TEs were also implicated in its relocation/distribution throughout the genome [31,54]. It is logical to believe that when tandem repeats are included within the mobile element sequence (for example, when the tandem repeats have its origin by duplication of part of TE sequence), maintaining the competence for mobilization for these TEs. The transposition mechanism can easily disperse tandem repeats throughout the genome (Figure 4A). This hypothesis is commonly accepted for short tandem repeats as micro and [115–117], which present short arrays of copies compared to satDNAs [39]. In fact, a considerable part of micro and minisatellites in eukaryotic genomes are embedded within mobile elements [115,117], which points to an important role of TEs in its genomic distribution, explaining its common dispersed chromosomal location. Moreover, TEs are present in pericentromeric regions of a wide range of species [27,70,74], being these regions also mainly built by satDNAs, which certainly facilitate the dispersion of these highly tandem repeats by retrotransposition.



**Figure 4.** Dispersion of tandem repeats by transposition. (A) Origin of tandem repeats from a part of a transposable element (TE) and its dispersion by transposition. The first duplications of a tandem repeat were originated from a part of a TE. As these repeats are included in the TE sequence, could then be dispersed by transposition. After, these repetitions may be amplified and homogenized in an array of copies. (B) Transduction of tandem repeats flanking a retrotransposon and its consequent dispersion throughout the genome by retrotransposition. Retrotransposon evidenced by a green block and tandem repeats evidenced by pink blocks. During the evolutionary time, the tandem repeats that were moved to new chromosomal locations could be amplified and homogenized, originating arrays of copies in these locations.

Regarding LINE-1 retrotransposons, several reports suggest a location of these elements in pericentromeric regions of different mammalian species chromosomes [27,66,70,118–120], pointing to an intermingling of these retrotransposons with satDNAs [66,118]. This complex organization pattern of repetitive sequences in the pericentromeric regions eventually favours the dispersion of satDNAs to other genomic locations by LINE-1 retrotransposition, since these elements frequently allow for the transduction of flanking non-LINE-1 DNA to new genomic locations (Figure 4B). This transduction is a consequence of a LINE-1 incorrect retrotransposition process [121–123]. Sometimes by retrotransposition, the TE sequences, along with its adjacent DNA, are copied and subsequently integrated into another genomic locations. This results in the duplication and genomic dispersion of the TE flanking sequences [124], as in, for example, satDNA monomers.

Furthermore, we can further speculate that DNA transposons can also allow the transduction of tandem repeats sequences [61], in a process similar to the one recognized in bacteria, for the transference of genes (e.g., antibiotic resistance genes) within and between bacterial genomes [125,126]. Sequences flanked by two “cut-and-paste” transposons can probably be mobilized in a genome, when its transposases use Terminal Inverted Repeats (TIRs) of the two different transposons to induce breaks for the mobilization (Figure 5A). If the TIRs of each DNA transposon are exclusively used, only these elements will be mobilized. Interestingly, as referred to previously, some similarity exists between the CENP-B box motifs and transposase recognition sites of DNA transposons [79], which may also lead to the identification of the CENP-B boxes as a break site for transposition [127]. Therefore, the common presence of CENP-B box motif in different satDNA families [127–130] can be involved in the mobilization of satDNAs copies during transposition [127]. No copies (monomers) of a satDNA described as presenting CENP-B boxes have these sequence motifs [131]. Thus, several monomers flanked by a DNA transposon and a CENP-box could be mobilized at the same time to another location, and subsequently amplified by different recombinational mechanisms (Figure 5B). This capacity of DNA transposons for the relocation of sequences flanked by them, or specifically by their TIRs, is indeed already used in medicine for gene therapy [132].



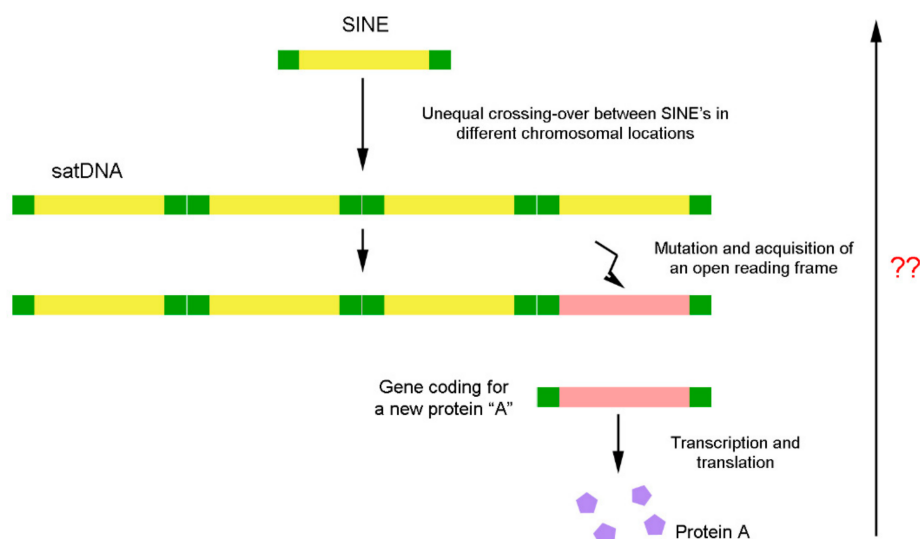
**Figure 5.** Dispersion of tandem repeats by “cut-and-paste” transposons. **(A)** Mobilization of sequences flanked by two “cut-and-paste” transposons. The breaks for the mobilization induced by transposases occur at the terminal inverted repeats (TIRs) of the two transposons. Yellow boxes: tandem repeats monomers, blue boxes: TIRs, violet boxes: transposase genes, grey boxes: remaining sequences of the chromosomes A and B. Chr: chromosome. **(B)** Mobilization of sequences flanked by a “cut-and-paste” transposon and a CENP-B box. The breaks for the mobilization occur at the TIRs of a transposon and a CENP-B box. Yellow boxes: tandem repeats monomers, blue boxes: TIRs, violet boxes: transposase genes, Orange box: CENP-B box, grey boxes: remaining sequences of the chromosomes.



### 3.2. Repetitive Sequences in the Origin of Coding Sequences

Transposable elements and satDNAs could also be involved in the evolution of genes, but most interestingly in the origin of new genes or gene variants. The noticeable ability of the TEs to produce genetic mutations when integrating at new genomic sites was recognized more than 50 years ago [20,133]. Nevertheless, despite most of these insertions being either neutral or deleterious to their host, its inclusions into new locations may also be advantageous, promoting gene evolution and the codification of more efficient protein variants. One of the most publicized discoveries about this subject is the resistance to HIV-1 (Human Immunodeficiency Virus) infection in owl monkeys, which presents an altered *TRIM5* gene with a cyclophilin A domain acquired by LINE-1 retrotransposition [109]. The binding of this cyclophilin domain to the HIV-1 viral capsid leads to a disruption of the infection process [134]. However, these primates are permissive to other immunodeficiency virus, such as the simian immunodeficiency virus (SIV) [109].

Recently, some works have shed light on questions about the de novo origin of protein-coding genes (or variants) from non-coding DNA [75,135–137], such as satDNAs. It is believed that the origin of completely novel genes from non-coding DNA is an evolutionary process comprising two big steps. In the first step, the non-coding DNA sequences are transcribed and then acquire translatable open reading frames [135] (Figure 6). Some works have already reported open reading frames in the monomers of satDNAs [138,139], an important step that might have allowed these sequences to evolve into coding sequences.



**Figure 6.** DNA remodelling process. Evolution of a satellite DNA sequence from a transposable element and its subsequent conversion in a coding sequence. The reverse sense of the process was not proved yet. ??- Is up to now unknown if occurs the opposite sense of this process for DNA sequences evolution.

## 4. Concluding Remarks

Pioneer studies on the eukaryotic genomic repetitive fraction has led to the classification of the repetitive sequences into two major groups according to the organization of copies within the genome: tandem repeats (as satDNAs) and dispersed repeats (TEs). Because of that, these sequences have been mostly investigated separately, with the important evolutionary relationships that exist between them not being considered. However, more precise genomic and bioinformatic analyses have now shown that these sequences do not have such a tight genomic organization. Some satDNAs may present dispersed isolated monomers in a genome [47] while TEs may have a kind of tandem organization of their copies [69]. Moreover, it became evident that a repetitive element can often be remodelled into a different sequence, a repetitive non-coding sequence or even a coding sequence. This suggests that the repetitive DNA elements in eukaryotic genomes seem to be in frequent remodulation, changing its organization

and function. Therefore, an update in repetitive sequence classification is now mandatory. Here, we have proposed a new classification for these sequences, not based on the genomic organization of their copies, but on other sequence features, namely the presence or absence of retrotransposon/transposon specific domains in their copies. Thus, we propose two new groups for the classification of the repetitive sequence: repeats presenting retrotransposons/transposons specific domains and repeats not presenting retrotransposons/transposons specific domains. This new classification demonstrates more clearly the evolutionary relationship between these sequences, promoting also more works to study together sequences that were previously considered very distinct. Their joint study is indeed very important for a better understanding of their function in genomes, showing that the evolutionary relationship between these sequences and the way that they can convert to each other is highly associated to the evolution of the genomes themselves.

Accordingly, we believe that future combined studies regarding TEs and tandem repeats, namely concerning chromosomal location and molecular similarity, will increase our knowledge about the evolution of eukaryotic genomes. The combined studies of related repetitive sequences can help us to understand the reason for some evolutionary tracks of sequences, and to understand these tracks in such a way that this genome plasticity makes the eukaryotic species better adapted to environmental conditions.

**Author Contributions:** Conceptualization, A.P., R.F. and A.V.-d.-S. and Y.Y.; writing—original draft preparation, A.P., R.F. and A.V.-d.-S.; writing—review and editing, A.P., R.F. and A.V.-d.-S.

**Funding:** This work was financially supported by a Newfelpro Post-doctoral grant (No. 82) from Republic of Croatia co-financed through the Marie Curie FP7-PEOPLE-2011-COFUND program.

**Acknowledgments:** We would like to acknowledge Raul Guizzo for carefully reviewing the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data and in the writing of the manuscript.

## References

1. Biscotti, M.A.; Olmo, E.; Heslop-Harrison, J.S. Repetitive DNA in eukaryotic genomes. *Chromosome Res.* **2015**, *23*, 415–420. [[CrossRef](#)] [[PubMed](#)]
2. Pucci, M.B.; Nogaroto, V.; Moreira-Filho, O.; Vicari, M.R. Dispersion of transposable elements and multigene families: Microstructural variation in *Characidium* (Characiformes: Crenuchidae) genomes. *Genet. Mol. Biol.* **2018**, *41*, 585–592. [[CrossRef](#)]
3. Venner, S.; Feschotte, C.; Biéumont, C. Dynamics of transposable elements: Towards a community ecology of the genome. *Trends Genet.* **2009**, *25*, 317–323. [[CrossRef](#)] [[PubMed](#)]
4. Devos, K.M. Grass genome organization and evolution. *Curr. Opin. Plant Biol.* **2010**, *13*, 139–145. [[CrossRef](#)] [[PubMed](#)]
5. Sun, C.; Shepard, D.B.; Chong, R.A.; Arriaza, J.L.; Hall, K.; Castoe, T.A.; Feschotte, C.; Pollock, D.D.; Mueller, R.L. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.* **2011**, *4*, 168–183. [[CrossRef](#)] [[PubMed](#)]
6. Lee, S.; Kim, N. Transposable elements and genome size variations in plants. *Genomics Inform.* **2014**, *12*, 87–97. [[CrossRef](#)]
7. Deniz, Ö.; Frost, J.M.; Branco, M.R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **2019**. [[CrossRef](#)]
8. Gregory, T.R. Animal genome size database. *Nucleic Acids Res.* **2007**, *35*, D332–D338. [[CrossRef](#)]
9. Shapiro, J.A.; von Sternberg, R. Why repetitive DNA is essential to genome function. *Biol. Rev. Camb. Philos. Soc.* **2005**, *80*, 227–250. [[CrossRef](#)]
10. Ferree, P.M.; Prasad, S. How can satellite DNA divergence cause reproductive Isolation? Let us count the chromosomal ways. *Genet. Res. Int.* **2012**, *2012*, 430136. [[CrossRef](#)]
11. Plohl, M.; Meštrović, N.; Mravinac, B. Satellite DNA Evolution. In *Repetitive DNA*; Garrido-Ramos, M.A., Ed.; Karger Genome Dyn: Basel, Switzerland, 2012; pp. 126–152.

12. Plohl, M.; Meštrović, N.; Mravinac, B. Centromere identity from the DNA point of view. *Chromosoma* **2014**, *123*, 313–325. [[CrossRef](#)] [[PubMed](#)]
13. Garrido-Ramos, M.A. Satellite DNA: An evolving topic. *Genes* **2017**, *8*, 230. [[CrossRef](#)] [[PubMed](#)]
14. Wichman, H.A.; Payne, C.T.; Ryder, O.A.; Hamilton, M.J.; Maltbie, M.; Baker, R.J. Genomic distribution of heterochromatic sequences in equids: Implications to rapid chromosomal evolution. *J. Hered.* **1991**, *82*, 369–377. [[CrossRef](#)] [[PubMed](#)]
15. Garagna, S.; Pérez-Zapata, A.; Zuccotti, M.; Mascheretti, S.; Marziliano, N.; Redi, C.A.; Aguilera, M.; Capanna, E. Genome composition in Venezuelan spiny-rats of the genus *Proechimys* (Rodentia, Echimyidae). I. Genome size, C-heterochromatin and repetitive DNAs in situ hybridization patterns. *Cytogenet. Cell Genet.* **1997**, *78*, 36–43. [[CrossRef](#)] [[PubMed](#)]
16. Feschotte, C.; Pritham, E.J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **2007**, *41*, 331–368. [[CrossRef](#)] [[PubMed](#)]
17. Vourc'h, C.; Biamonti, G. Long non-coding RNAs. Progress in Molecular and Subcellular Biology. In *Transcription of Satellite DNAs in Mammals*; Ugarković, D., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 95–118.
18. Zhu, Q.; Pao, G.M.; Huynh, A.M.; Suh, H.; Tonnu, N.; Nederlof, P.M.; Gage, F.H.; Verma, I.M. *BRCA1* tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **2011**, *477*, 179–184. [[CrossRef](#)] [[PubMed](#)]
19. Hall, L.E.; Mitchell, S.E.; O'Neill, R.J. Pericentric and centromeric transcription: A perfect balance required. *Chromosome Res.* **2012**, *20*, 535–546. [[CrossRef](#)]
20. Rebollo, R.; Romanish, M.T.; Mager, D.L. Transposable elements: An abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **2012**, *46*, 21–42. [[CrossRef](#)]
21. Erukashvily, N.I.; Ponomartsev, N.V. Mammalian satellite DNA: A speaking dumb. *Adv. Protein Chem. Struct. Biol.* **2013**, *90*, 31–65.
22. Belyayev, A. Bursts of transposable elements as an evolutionary driving force. *J. Evol. Biol.* **2014**, *27*, 2573–2584. [[CrossRef](#)]
23. Cridland, J.M.; Thornton, K.R.; Long, A.D. Gene expression variation in *Drosophila melanogaster* due to rare transposable element insertion alleles of large effect. *Genetics* **2014**, *199*, 85–93. [[CrossRef](#)] [[PubMed](#)]
24. Friedli, M.; Trono, D. The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* **2015**, *31*, 429–451. [[CrossRef](#)] [[PubMed](#)]
25. Kim, Y.J.; Han, K. Endogenous retrovirus-mediated genomic variations in chimpanzees. *Mob. Genet. Elements* **2015**, *4*, 1–4. [[CrossRef](#)] [[PubMed](#)]
26. Paço, A.; Adegá, F.; Chaves, R. Line-1 retrotransposons: From “parasite” sequences to functional elements. *J. Appl. Genet.* **2015**, *56*, 133–145. [[CrossRef](#)]
27. Paço, A.; Adegá, F.; Meštrović, N.; Plohl, M.; Chaves, R. The puzzling character of repetitive DNA in *Phodopus* genomes (Cricetidae, Rodentia). *Chromosome Res.* **2015**, *23*, 427–440. [[CrossRef](#)]
28. Feliciello, I.; Akrap, I.; Ugarković, D. Satellite DNA modulates gene expression in the beetle *Tribolium castaneum* after heat stress. *PLoS Genet.* **2015**, *11*, e1005547.
29. Song, J.; Dong, F.; Lilly, J.W.; Stupar, R.M.; Jiang, J. Instability of bacterial artificial chromosome (BAC) clones containing tandemly repeated DNA sequences. *Genome* **2001**, *44*, 463–469. [[CrossRef](#)]
30. Alkan, C.; Kidd, J.M.; Marques-Bonet, T.; Aksay, G.; Antonacci, F.; Hormozdiari, F.; Kitzman, J.O.; Baker, C.; Malig, M.; Mutlu, O.; et al. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat. Genet.* **2009**, *41*, 1061–1067. [[CrossRef](#)]
31. Macas, J.; Koblížková, A.; Navrátilová, A.; Neumann, P. Hypervariable 3' UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. *Gene* **2009**, *448*, 198–206. [[CrossRef](#)]
32. Kass, D.H.; Batzer, M.A. *Genome Organization/Human*; ELS: Princeton, NJ, USA, 2001; pp. 1–8.
33. Richard, G.F.; Kerrest, A.; Dujon, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 686–727. [[CrossRef](#)]
34. Petrov, D.A. Evolution of genome size: New approaches to an old problem. *Trends Genet.* **2001**, *17*, 23–28. [[CrossRef](#)]
35. Boulesteix, M.; Weiss, M.; Biéumont, C. Differences in the genome size between closely related species: The *Drosophila melanogaster* species subgroup. *Mol. Biol. Evol.* **2006**, *23*, 162–167. [[CrossRef](#)]
36. Pritham, E.J. Transposable elements and factors influencing their success in eukaryotes. *J. Hered.* **2009**, *100*, 648–655. [[CrossRef](#)]

37. Wheeler, T.J.; Clements, J.; Eddy, S.R.; Hubley, R.; Jones, T.A.; Jurka, J.; Smit, A.F.A.; Finn, R.D. Dfam: A database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res.* **2013**, *41*, D70–D82. [[CrossRef](#)]
38. Hubley, R.; Finn, R.D.; Clements, J.; Eddy, S.R.; Jones, T.A.; Bao, W.; Smit, A.F.; Wheeler, T.J. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **2016**, *44*, D81–D89. [[CrossRef](#)]
39. Slamovits, C.H.; Rossi, M.S. Satellite DNA: Agent of chromosomal evolution in mammals. A review. *J. Neotrop. Mamm.* **2002**, *9*, 297–308.
40. Jurka, J.; Kapitonov, V.V.; Kohany, O.; Jurka, M.V. Repetitive sequences in complex genomes: Structure and evolution. *Annu. Rev. Genomics Hum. Genet.* **2007**, *8*, 241–259. [[CrossRef](#)]
41. Kapitonov, V.V.; Jurka, J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* **2008**, *9*, 411–412. [[CrossRef](#)]
42. Kapitonov, V.V.; Tempel, S.; Jurka, J. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domains protein sequence. *Gene* **2009**, *448*, 207–213. [[CrossRef](#)]
43. Strachan, T.; Read, A.P. *Human Molecular Genetics 3*; Garland Science: London, UK, 2004.
44. Ruiz-Ruano, F.J.; López-León, M.D.; Cabrero, J.; Camacho, J.P. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* **2016**, *6*, 28333. [[CrossRef](#)]
45. Schlötterer, C.; Harr, B. *Microsatellite Instability*; ELS: Princeton, NJ, USA, 2001; pp. 1–4.
46. Subirana, J.A.; Messegue, X. Evolution of tandem repeat satellite sequences in two closely related *Caenorhabditis* species. Diminution of Satellites in Hermaphrodites. *Genes* **2017**, *8*, 351. [[CrossRef](#)]
47. Louzada, S.; Vieira-da-Silva, A.; Mendes-da-Silva, A.; Kubickova, S.; Rubes, J.; Adegas, 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. [[CrossRef](#)]
48. Paço, A.; Adegas, F.; Meštrović, N.; Plohl, M.; Chaves, R. Evolutionary story of a satellite DNA from *Phodopus sungorus* (Rodentia, Cricetidae). *Genome Biol. Evol.* **2014**, *6*, 2944–2955. [[CrossRef](#)]
49. Frommer, M.; Prosser, J.; Tkachuk, D.; Reisner, A.H.; Vincent, P.C. Simple repeated sequences in human satellite DNA. *Nucleic Acids Res.* **1982**, *10*, 547–563. [[CrossRef](#)]
50. Modi, W.S. Rapid, localized amplification of a unique satellite DNA family in the rodent *Microtus chrotorrhinus*. *Chromosoma* **1993**, *102*, 484–490. [[CrossRef](#)] [[PubMed](#)]
51. Schmidt, T.; Heslop-Harrison, J.S. Genomes, genes and junk: The largescale organization of plant chromosomes. *Trends Plant. Sci.* **1998**, *3*, 195–199. [[CrossRef](#)]
52. Henikoff, S.; Ahmad, K.; Malik, H.S. The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* **2001**, *293*, 1098–1102. [[CrossRef](#)]
53. Li, W.H. *Molecular Evolution*; Sinauer: Sunderland, MA, USA, 1997.
54. Palomeque, T.; Lorite, P. Satellite DNA in insects: A review. *Heredity* **2008**, *100*, 564–573. [[CrossRef](#)]
55. Brajković, J.; Feliciello, I.; Bruvo-Madarić, B.; Ugarković, D. Satellite DNA-like elements associated with genes within euchromatin of the beetle *Tribolium Castaneum*. *G3* **2012**, *2*, 931–941. [[CrossRef](#)]
56. Vittorazzi, S.E.; Lourenço, L.B.; Recco-Pimentel, S.M. Long-time evolution and highly dynamic satellite DNA in leptodactylid and hylodid frogs. *BMC Genet.* **2014**, *15*, 111. [[CrossRef](#)]
57. Escudeiro, A.; Adegas, F.; Robinson, T.J.; Heslop-Harrison, J.S.; Chaves, R. Conservation, divergence, and functions of centromeric satellite DNA families in the Bovidae. *Genome Biol. Evol.* **2019**, *11*, 1152–1165. [[CrossRef](#)] [[PubMed](#)]
58. Kronmiller, B.A.; Wise, R.P. TEnest: Automated chronological annotation and visualization of nested plant transposable elements. *Plant Physiol.* **2008**, *146*, 45–59. [[CrossRef](#)] [[PubMed](#)]
59. Goerner-Potvin, P.; Bourque, G. Computational tools to unmask transposable elements. *Nat. Rev. Genet.* **2018**, *19*, 688–704. [[CrossRef](#)] [[PubMed](#)]
60. Bourque, G.; Burns, K.H.; Gehring, M.; Gorbunova, V.; Seluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H.L.; Macfarlan, T.S.; et al. Ten things you should know about transposable elements. *Genome Biol.* **2018**, *19*, 199. [[CrossRef](#)]
61. Arkhipova, I.R.; Yushenova, I.A. Giant transposons in eukaryotes: Is bigger better? *Genome Biol. Evol.* **2019**, *11*, 906–918. [[CrossRef](#)]
62. Grahn, R.A.; Rinehart, T.A.; Cantrell, M.A.; Wichman, H.A. Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents. *Cytogenet. Genome Res.* **2005**, *110*, 407–415. [[CrossRef](#)]

63. Konkel, M.K.; Batzer, M.A. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Semin. Cancer Biol.* **2010**, *20*, 211–221. [[CrossRef](#)]
64. Soares, M.L.; Edwards, C.A.; Dearden, F.L.; Ferrón, S.R.; Curran, S.; Corish, J.A.; Rancourt, R.C.; Allen, S.E.; Charalambous, M.; Ferguson-Smith, M.A.; et al. Targeted deletion of a 170-kb cluster of LINE-1 repeats and implications for regional control. *Genome Res.* **2018**. [[CrossRef](#)]
65. Dobigny, G.; Ozouf-Costaz, C.; Bonillo, C.; Volobouev, V. Viability of X-autosome translocations in mammals: An epigenomic hypothesis from a rodent case-study. *Chromosoma* **2004**, *113*, 34–41. [[CrossRef](#)]
66. Marchal, J.A.; Acosta, M.J.; Bullejos, M.; Puerma, E.; Díaz de la Guardia, R.; Sánchez, A. Distribution of L1-retroposons on giant sex chromosomes of *Microtus cabrerai* (Arvicolidae, Rodentia): Functional and evolutionary implications. *Chromosome Res.* **2006**, *14*, 177–186. [[CrossRef](#)]
67. Giordano, J.; Ge, Y.; Gelfand, Y.; Abrusán, G.; Benson, G.; Warburton, P.E. Evolutionary history of mammalian transposons determined by genome-wide defragmentation. *PLoS Comput. Biol.* **2007**, *3*, e137. [[CrossRef](#)] [[PubMed](#)]
68. Rebuzzini, P.; Castiglia, R.; Nergadze, S.G.; Mitsainas, G.; Munclinger, P.; Zuccotti, M.; Capanna, E.; Redi, C.A.; Garagna, S. Quantitative variation of LINE-1 sequences in five species and three subspecies of the subgenus *Mus* and in five Robertsonian races of *Mus musculus domesticus*. *Chromosome Res.* **2009**, *17*, 65–76. [[CrossRef](#)]
69. Dhillon, B.; Gill, N.; Hamelin, R.C.; Goodwin, S.B. The landscape of transposable elements in the finished genome of the fungal wheat pathogen *Mycosphaerella graminicola*. *BMC Genomics* **2014**, *15*, 1132. [[CrossRef](#)] [[PubMed](#)]
70. Vieira-da-Silva, A.; Adegá, F.; Guedes-Pinto, H.; Chaves, R. LINE-1 distribution in six rodent genomes follow a species-specific pattern. *J. Genet.* **2016**, *95*, 21–33. [[CrossRef](#)]
71. Slijepcevic, P. Telomeres and mechanisms of Robertsonian fusion. *Chromosoma* **1998**, *107*, 136–140. [[CrossRef](#)]
72. Bolzán, A.D.; Bianchi, M.S. Telomeres, interstitial telomeric repeat sequences, and chromosomal aberrations. *Mutat. Res.* **2006**, *612*, 189–214. [[CrossRef](#)]
73. Paço, A.; Chaves, R.; Vieira-da-Silva, A.; Adegá, F. The involvement of repetitive sequences in the remodelling of karyotypes: The Phodopus genomes (Rodentia, Cricetidae). *Micron* **2013**, *46*, 27–34. [[CrossRef](#)]
74. Wong, L.H.; Choo, K.H.A. Evolutionary dynamics of transposable elements at the centromere. *Trends Genet.* **2004**, *20*, 611–616. [[CrossRef](#)]
75. Xie, C.; Zhang, Y.E.; Chen, J.Y.; Liu, C.J.; Zhou, W.Z.; Li, Y.; Zhang, M.; Zhang, R.; Wei, L.; Li, C.Y. Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs. *PLoS Genet.* **2012**, *8*, e1002942. [[CrossRef](#)]
76. Kapitonov, V.V.; Holmquist, G.P.; Jurka, J. L1 repeat is a basic unit of heterochromatin satellites in cetaceans. *Mol. Biol. Evol.* **1998**, *15*, 611–612. [[CrossRef](#)]
77. Plohl, M.; Petrović, V.; Luchetti, A.; Ricci, A.; Šatović, E.; Passamonti, M.; Mantovani, B. Long-term conservation vs. high sequence divergence: The case of an extraordinarily old satellite DNA in bivalve molluscs. *Heredity* **2010**, *104*, 543–551. [[CrossRef](#)] [[PubMed](#)]
78. Meštrović, N.; Mravinac, B.; Pavlek, M.; Vojvoda-Zeljko, T.; Šatović, E.; Plohl, M. Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosome Res.* **2015**, *23*, 583–596. [[CrossRef](#)] [[PubMed](#)]
79. Kipling, D.; Warburton, P.E. Centromeres, CENP-B and Tigger too. *Trends Genet.* **1997**, *13*, 141–145. [[CrossRef](#)]
80. Smith, G.P. Evolution of repeated DNA sequences by unequal crossover. *Science* **1976**, *191*, 528–535. [[CrossRef](#)] [[PubMed](#)]
81. Rossi, M.S.; Pesce, C.G.; Reig, O.A.; Kornblihtt, A.R.; Zorzópulos, J. Retroviral-like features in the repetitive unit of the major satellite DNA from the South American rodents of the genus *Ctenomys*. *DNA Seq.* **1993**, *3*, 379–381. [[CrossRef](#)] [[PubMed](#)]
82. Batistoni, R.; Pesole, G.; Marracci, S.; Nardi, I. A tandemly repeated DNA family originated from SINE-related elements in the European *Plethodontid Salamanders* (Amphibia, Urodela). *J. Mol. Evol.* **1995**, *40*, 608–615. [[CrossRef](#)]
83. Heikkinen, E.; Launonen, V.; Müller, E.; Bachmann, L. The pvB370 BamHI satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic pDv elements. *J. Mol. Evol.* **1995**, *41*, 604–614. [[CrossRef](#)]
84. Kapitonov, V.V.; Jurka, J. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* **1999**, *107*, 27–37. [[CrossRef](#)]

85. López-Flores, I.; de la Herrán, R.; Garrido-Ramos, M.A.; Boudry, P.; Ruiz-Rejón, C.; Ruiz-Rejón, M. The molecular phylogeny of oysters based on a satellite DNA related to transposons. *Gene* **2004**, *339*, 181–188. [[CrossRef](#)]
86. Coates, B.S.; Kroemer, J.A.; Sumerford, D.V.; Hellmich, R.L. A novel class of miniature inverted repeat transposable elements (MITEs) that contain hitchhiking (GTCY)<sub>n</sub> microsatellites. *Insect Mol. Biol.* **2011**, *20*, 15–27. [[CrossRef](#)]
87. Sharma, A.; Wolfgruber, T.K.; Presting, G.G. Tandem repeats derived from centromeric retrotransposons. *BMC Genomics* **2013**, *14*, 1–11. [[CrossRef](#)] [[PubMed](#)]
88. Dias, G.B.; Heringer, P.; Svartman, M.; Kuhn, G.C.S. Helitrons shaping the genomic architecture of *Drosophila*: Enrichment of DINE-TR1 in  $\alpha$ - and  $\beta$ -heterochromatin, satellite DNA emergence, and piRNA expression. *Chromosome Res.* **2015**, *23*, 597–613. [[CrossRef](#)] [[PubMed](#)]
89. Pasero, P.; Sjakste, N.; Blettry, C.; Got, C.; Marilley, M. Long-range organization and sequence-directed curvature of *Xenopus laevis* satellite 1 DNA. *Nucleic Acids Res.* **1993**, *21*, 4703–4710. [[CrossRef](#)] [[PubMed](#)]
90. Agudo, M.; Losada, A.; Abad, J.P.; Pimpinelli, S.; Ripoll, P.; Villasante, A. Centromeres from telomeres? The centromeric region of the Y chromosome of *Drosophila melanogaster* contains a tandem array of telomeric HeT-A- and TART-related sequences. *Nucleic Acids Res.* **1999**, *27*, 3318–3324. [[CrossRef](#)] [[PubMed](#)]
91. Langdon, T.; Seago, C.; Jones, R.N.; Ougham, H.; Thomas, H.; Forster, J.W.; Jenkins, G. De novo evolution of satellite DNA on the rye B chromosome. *Genetics* **2000**, *154*, 869–884. [[PubMed](#)]
92. Miller, W.J.; Nagel, A.; Bachmann, J.; Bachmann, L. Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. *Mol. Biol. Evol.* **2000**, *17*, 1597–1609. [[CrossRef](#)]
93. Cheng, Z.J.; Murata, M. A centromeric tandem repeat family originating from a part of Ty3/gypsy-retroelement in wheat and its relatives. *Genetics* **2003**, *164*, 665–672.
94. Hikosaka, A.; Kawahara, A. Lineage-specific tandem repeats riding on a transposable element of MITE in *Xenopus* evolution: A new mechanism for creating simple sequence repeats. *J. Mol. Evol.* **2004**, *59*, 738–746. [[CrossRef](#)]
95. López-Flores, I.; Garrido-Ramos, M.A. The repetitive DNA content of Eucaryotic genomes. In *Repetitive DNA*; Garrido-Ramos, M.A., Ed.; Karger Genome Dyn: Basel, Switzerland, 2012; pp. 1–28.
96. Tek, A.L.; Song, J.; Macas, J.; Jiang, J. Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. *Genetics* **2005**, *170*, 1231–1238. [[CrossRef](#)]
97. Li, J.; Leung, F.C. A CR1 element is embedded in a novel tandem repeat (*HinfI* repeat) within the chicken genome. *Genome* **2006**, *49*, 97–103. [[CrossRef](#)]
98. 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](#)] [[PubMed](#)]
99. Arcot, S.S.; Wang, Z.; Weber, J.L.; Deininger, P.L.; Batzer, M.A. Alu repeats: A source for the genesis of primate microsatellites. *Genomics* **1995**, *29*, 136–144. [[CrossRef](#)] [[PubMed](#)]
100. Ramsay, L.; Macaulay, M.; Cardle, L.; Morgante, M.; degli Ivanissevich, S.; Maestri, E.; Powell, W.; Waugh, R. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.* **1999**, *17*, 415–425. [[CrossRef](#)] [[PubMed](#)]
101. Duffy, A.J.; Coltman, D.W.; Wright, J.M. Microsatellites at a common site in the second ORF of L1 elements in mammalian genomes. *Mamm. Genome* **1996**, *7*, 386–387. [[CrossRef](#)] [[PubMed](#)]
102. Clark, R.M.; Dalglish, G.L.; Endres, D.; Gomez, M.; Taylor, J.; Bidichandani, S.I. Expansion of GAA triplet repeats in the human genome: Unique origin of the FRDA mutation at the center of an *Alu*. *Genomics* **2004**, *83*, 373–383. [[CrossRef](#)]
103. Armour, J.A.L.; Wong, Z.; Wilson, V.; Royle, N.J.; Jeffreys, A.J. Sequences flanking the repeat arrays of human minisatellites: Association with tandem and dispersed repeat elements. *Nucleic Acids Res.* **1989**, *17*, 4925–4935. [[CrossRef](#)]
104. Kelly, R.G. Similar origins of two mouse minisatellites within transposon-like LTRs. *Genomics* **1994**, *24*, 509–515. [[CrossRef](#)]
105. Bois, P.; Williamson, J.; Brown, J.; Dubrova, Y.E.; Jeffreys, A.J. A novel unstable mouse VNTR family expanded from SINE B1 elements. *Genomics* **1998**, *49*, 122–128. [[CrossRef](#)]
106. Jurka, J.; Gentles, A.J. Origin and diversification of minisatellites derived from human Alu sequences. *Gene* **2006**, *365*, 21–26. [[CrossRef](#)]

107. Ahmed, M.; Liang, P. Transposable elements are a significant contributor to tandem repeats in the human genome. *Comp. Funct. Genomics* **2012**, *2012*, 947089. [[CrossRef](#)]
108. Mogil, L.S.; Slowikowski, K.; Laten, H.M. Computational and experimental analyses of retrotransposon-associated minisatellite DNAs in the soybean genome. *BMC Bioinform.* **2012**, *13* (Suppl. 2), S13. [[CrossRef](#)] [[PubMed](#)]
109. Sayah, D.M.; Sokolskaja, E.; Berthoux, L.; Luban, J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **2004**, *430*, 569–573. [[CrossRef](#)] [[PubMed](#)]
110. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [[CrossRef](#)] [[PubMed](#)]
111. Waterston, R.H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J.F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; Mouse genome sequencing consortium; et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **2002**, *420*, 520–562.
112. Levinson, G.; Gutman, G.A. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **1987**, *4*, 203–221.
113. Buschiazzo, E.; Gemmell, N.J. The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays* **2006**, *28*, 1040–1050. [[CrossRef](#)]
114. Haber, J.E.; Louis, E.J. Minisatellite origins in yeast and humans. *Genomics* **1998**, *48*, 132–135. [[CrossRef](#)]
115. Inukai, T. Role of transposable elements in the propagation of minisatellites in the rice genome. *Mol. Genet. Genomics* **2004**, *271*, 220–227. [[CrossRef](#)]
116. Van't Hof, A.E.; Brakefield, P.M.; Saccheri, I.J.; Zwaan, B.J. Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. *Heredity* **2007**, *98*, 320–328. [[CrossRef](#)]
117. Smýkal, P.; Kalendar, R.; Ford, R.; Macas, J.; Griga, M. Evolutionary conserved lineage of Angela-family retrotransposons as a genome-wide microsatellite repeat dispersal agent. *Heredity* **2009**, *103*, 157–167. [[CrossRef](#)]
118. Mayorov, V.I.; Adkison, L.R.; Vorobyeva, N.V.; Khrapov, E.A.; Kholodhov, N.G.; Rogozin, I.B.; Nesterova, T.B.; Protopopov, A.I.; Sablina, O.V.; Graphodatsky, A.S.; et al. Organization and chromosomal localization of a B1-like containing repeat of *Microtus subarvalis*. *Mamm. Genome* **1996**, *7*, 593–597. [[CrossRef](#)] [[PubMed](#)]
119. Waters, P.D.; Dobigny, G.; Pardini, A.T.; Robinson, T.J. LINE-1 distribution in Afrotheria and Xenarthra: Implications for understanding the evolution of LINE-1 in eutherian genomes. *Chromosoma* **2004**, *113*, 137–144. [[CrossRef](#)] [[PubMed](#)]
120. Acosta, M.J.; Marchal, J.A.; Fernández-Espartero, C.H.; Bullejos, M.; Sánchez, A. Retroelements (LINEs and SINEs) in vole genomes: Differential distribution in the constitutive heterochromatin. *Chromosome Res.* **2008**, *16*, 949–959. [[CrossRef](#)] [[PubMed](#)]
121. Goodier, J.L.; Ostertag, E.M.; Du, K.; Kazazian, H.H., Jr. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res.* **2001**, *11*, 1677–1685. [[CrossRef](#)] [[PubMed](#)]
122. Babushok, D.V.; Ohshima, K.; Ostertag, E.M.; Chen, X.; Wang, Y.; Mandal, P.K.; Okada, N.; Abrams, C.S.; Kazazian, H.H., Jr. A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. *Genome Res.* **2007**, *17*, 1129–1138. [[CrossRef](#)] [[PubMed](#)]
123. Cordaux, R.; Batzer, M.A. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **2009**, *10*, 691–703. [[CrossRef](#)]
124. Ewing, A.D. Transposable element detection from whole genome sequence data. *Mob. DNA* **2015**, *6*, 24. [[CrossRef](#)]
125. Ochman, H.; Lawrence, J.G.; Groisman, E.A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **2000**, *405*, 299–304. [[CrossRef](#)]
126. Sun, W.; Gu, J.; Wang, X.; Qian, X.; Tuo, X. Impacts of biochar on the environmental risk of antibiotic resistance genes and mobile genetic elements during anaerobic digestion of cattle farm wastewater. *Bioresour. Technol.* **2018**, *256*, 342–349. [[CrossRef](#)]
127. Meštrović, N.; Pavlek, M.; Car, A.; Castagnone-Sereno, P.; Abad, P.; Plohl, M. Conserved DNA motifs, including the CENP-B Box-like, are possible promoters of satellite DNA array rearrangements in nematodes. *PLoS ONE* **2013**, *8*, e67328. [[CrossRef](#)]
128. Canapa, A.; Barucca, M.; Cerioni, P.N.; Olmo, E. A satellite DNA containing CENP-B box-like motifs is present in the Antarctic scallop *Adamussium colbecki*. *Gene* **2000**, *247*, 175–180. [[CrossRef](#)]

129. Lorite, P.; Carrillo, J.A.; Tinaut, A.; Palomeque, T. Evolutionary dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera, Formicidae). *Gene* **2004**, *332*, 159–168. [[CrossRef](#)] [[PubMed](#)]
130. Mravinac, B.; Ugarković, Đ.; Franjević, D.; Plohl, M. Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. *J. Mol. Evol.* **2005**, *60*, 513–525. [[CrossRef](#)] [[PubMed](#)]
131. 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 centromere satellite. *J. Cell Biol.* **1989**, *109*, 1963–1973. [[CrossRef](#)]
132. Rajabpour, F.V.; Raoofian, R.; Habibi, L.; Akrami, S.M.; Tabrizi, M. Novel trends in genetics: Transposable elements and their application in medicine. *Arch. Iran. Med.* **2014**, *17*, 702–712.
133. Dubin, M.J.; Mittelsten Scheid, O.; Becker, C. Transposons: A blessing curse. *Curr. Opin. Plant Biol.* **2018**, *42*, 23–29. [[CrossRef](#)]
134. Frausto, S.D.; Lee, E.; Tang, H. Cyclophilins as modulators of viral replication. *Viruses* **2013**, *5*, 1684–1701. [[CrossRef](#)]
135. Wu, D.D.; Irwin, D.M.; Zhang, Y.P. De novo origin of human protein-coding genes. *PLoS Genet.* **2011**, *7*, e1002379. [[CrossRef](#)]
136. Murphy, D.N.; McLysaght, A. De novo origin of protein-coding genes in murine rodents. *PLoS ONE* **2012**, *7*, e48650. [[CrossRef](#)]
137. Wu, B.; Knudson, A. Tracing the de novo origin of protein-coding genes in yeast. *MBio* **2018**, *9*, e01024-18. [[CrossRef](#)]
138. Schmidt, T.; Metzloff, M. Cloning and characterization of a *Beta vulgaris* satellite DNA family. *Gene* **1991**, *101*, 247–250. [[CrossRef](#)]
139. Passamonti, M.; Mantovani, B.; Scali, V. Characterization of a highly repeated DNA family in *Tapetinae* Species (Mollusca Bivalvia: Veneridae). *Zool. Sci* **1998**, *15*, 599–605. [[CrossRef](#)]



© 2019 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/>).