Review

On the Base Composition of Transposable Elements

Stéphane Boissinot [ID]

Center for Genomics and Systems Biology, New York University Abu Dhabi, Saadiyat Island,
Abu Dhabi P.O. Box 129188, United Arab Emirates; sb5272@nyu.edu

Abstract: Transposable elements exhibit a base composition that is often different from the genomic average and from hosts' genes. The most common compositional bias is towards Adenosine and Thymine, although this bias is not universal, and elements with drastically different base composition can coexist within the same genome. The AT-richness of transposable elements is apparently maladaptive because it results in poor transcription and sub-optimal translation of proteins encoded by the elements. The cause(s) of this unusual base composition remain unclear and have yet to be investigated. Here, I review what is known about the nucleotide content of transposable elements and how this content can affect the genome of their host as well as their own replication. The compositional bias of transposable elements could result from several non-exclusive processes including horizontal transfer, mutational bias, and selection. It appears that mutation alone cannot explain the high AT-content of transposons and that selection plays a major role in the evolution of the compositional bias. The reason why selection would favor a maladaptive nucleotide content remains however unexplained and is an area of investigation that clearly deserves attention.

Keywords: transposable elements; GC content; base composition; codon bias

1. Introduction

The base composition is one of the most fundamental properties of a genome or of a DNA sequence. Although all DNA sequences consist of 4 nucleotides, the relative proportion of Guanine/Cytosine (GC%) and Adenosine/Thymine (AT%) can differ considerably among organisms and among genomic regions. For instance, mammalian and avian genomes are highly heterogenous in base content, with gene-rich GC-rich compartments embedded in AT-rich intergenic regions, while reptiles, amphibians and fish are generally homogenous in base composition [1,2]. At a smaller scale, the GC% may vary among genes, among regions of a gene but also among codon positions. These differences in base composition can in turn affect a number of fundamental biological processes including transcription efficacy [3,4], the secondary structure of RNA molecules, translation efficacy and accuracy [5–8], the amino acid composition of proteins, and epigenetic modifications of DNA.

Transposable elements (TEs) are major components of genomes and have a profound impact on the size, structure, and function of their hosts' genomes (Reviewed in [9]). Although most TE insertions are neutral or deleterious, TEs can also be a source of new genes or of regulatory motifs [9–12]. An aspect that has received little attention is the impact TEs can have on their host's genome in terms of base composition. In many organisms, the base composition of TEs differ drastically from the genomic average and from hosts' genes [13–17], to the point that the unusual base composition of TEs can be used to detect them in genomes [18]. This compositional bias, which is most commonly an AT-bias, may thus impact the structure and function of the genome in a number of ways. For instance, the accumulation of a type of TEs in specific genomic regions can potentially affect the GC genomic landscape, which in turn can affect other biological properties such as chromatin structure. Another interesting aspect is the effect the base composition of TEs can have on their own replication. In a number of organisms, the high AT% exhibited by TEs results

in poor transcription and sub-optimal translation of TE-encoded proteins and thus seems maladaptive. Nevertheless, the AT-richness of TEs is widespread, and the persistence of such an unusual base composition across many categories of TEs remains a puzzle.

Here, I will review the state of our knowledge on the evolution of base composition in TEs, as well as the numerous questions that remain unanswered on this topic. After a short introduction on the biology of TEs, I will review what is known about their base composition, and in particular, I will emphasize that the high AT content observed for many TEs in many organisms is, in fact, not universal. I will then describe the consequences the unusual base composition of TEs may have on their hosts but also on their own replication. I will finally explore the evolutionary processes that are potentially driving the base composition of TEs towards nucleotide contents that appear, at first, maladaptive.
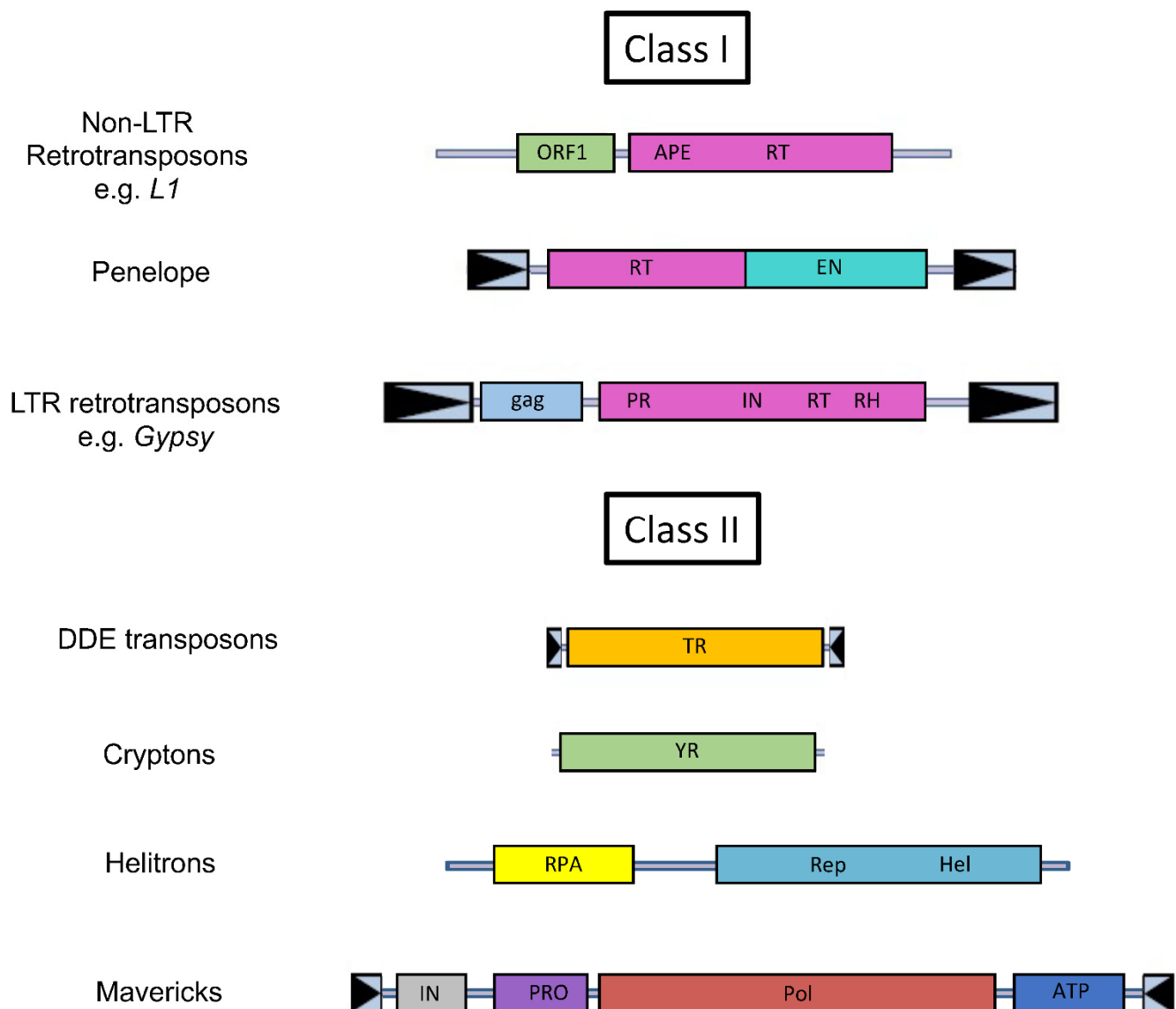
2. A Primer on Transposable Elements

Transposable elements constitute a diverse group of sequences that have in common the ability to move from one location in the genome of their host to another location [19,20]. They are typically classified based on their mode of mobility [21,22]. Elements that move using an RNA intermediate are called class I elements, and those that do not are called class II (Figure 1). Each of these classes contains a myriad of subsets. Class I elements are further divided into LTR-retrotransposons, which are flanked by Long Terminal Repeats (LTRs) and include LTR-retrotransposons sensus stricto, endogenous retroviruses and *DIRS* elements, non-LTR retrotransposons (also called Long Interspersed Nuclear Elements or *LINEs*), and *Penelope* Elements. All these elements have in common the use of a reverse-transcriptase for their replication [23]. The replicative machinery of class I elements can also act on other transcripts and is responsible for the amplification of non-autonomous retroelements (such as Short INterspersed Elements or *SINEs*), which can far outnumber their autonomous progenitors [24–26]. Class II elements constitute a very disparate group of elements [27], which only have in common the fact that their replication does not require an RNA intermediate. They consist of four subgroups: DDE transposons that mobilized by a cut-and-paste mechanism mediated by a transposase, *Cryptons* that use a tyrosine recombinase for their transposition, *Helitrons* that use a rolling-circle mode of replication [28,29], and *Mavericks* that are mobilized by a self-synthetizing process mediated by a protein-primed polymerase B [30,31]. Class II elements can also mediate the transposition of non-autonomous copies, which can outnumber autonomous copies [32–34].

Class I and class II elements also differ by their long-term evolutionary dynamics within their host. Most *LINEs* in vertebrates are transmitted vertically over extended periods of evolutionary time and are thus long-term residents of these genomes. For instance, *L1* retrotransposons have persisted in the genome of mammals since the origin of this vertebrate class, and mammalian genomes contain a near complete record of the successive waves of *L1* amplification they have experienced since their origin [35,36]. The investigation of *L1* in mammals revealed that a very small number of lineages, often only one, persisted over long periods of time [35,37–40], which could reflect an arms race between *L1* and the repression machinery of the host [41,42]. In contrast, class II elements tend to invade the genome of their hosts by horizontal transfer, then amplify to large number but eventually get extinct [43–45]. Consequently, they rarely persist for long periods of time and are typically transient residents of genomes.

The number of TE copies and the diversity of elements differ considerably among genomes and depends on a number of parameters including the rate of transposition, the rate of fixation and the rate of DNA loss caused by deletions [46]. The rate of transposition depends on the number of progenitor copies and on the location of these progenitors in the genome (transcriptionally active vs. inactive genomic regions). The rate of transposition will also be affected by host-encoded repression processes. Since TE activity can be deleterious, a number of defense mechanisms have evolved to protect the integrity of the genome [42,47,48], DNA methylation being the best-known mechanism of defense against TE activity [49,50]. The rate of fixation will depend on the combined effect of purifying

selection and genetic drift (reviewed in [46]) as well as linked selection [51]. Since the majority of new insertions is either deleterious or neutral, most of them are not expected to remain in the population and to be lost by chance (in the case of a neutral insertion) or to be eliminated by purifying selection (if the insertion is deleterious) [52–55]. However, in small populations, genetic drift can counteract the effect of selection and deleterious insertions can reach fixation [51,56–59]. Thus, one can expect that the overall rate of fixation and the accumulation of new copies will be higher in small populations than in large populations [60]. Finally, the number of TE derived sequence in a genome will depend on the rate of decay of these elements resulting from the rate of DNA loss by large deletions, which was shown to differ among organisms [33,61,62] and may be correlated to the number of copies (i.e., the accordion model of evolution) [63].

Figure 1. Schematic representation of the main categories of autonomous transposable elements. The elements are not drawn to scale. The following abbreviations are used: APE, apurinic endonuclease; RT, reverse transcriptase; ORF1, open-reading frame 1; EN, GIY-YIG endonuclease; gag, gag gene; PR, proteinase; IN, integrase; RH, RNase H domain; TR, transposase; YR, tyrosine recombinase; RPA, replication protein A; Rep, replication initiation domain; Hel, helicase; PRO, cysteine protease; Pol, protein-primed type B DNA polymerase; ATP, ATPase. The boxed arrows represent terminal repeats.

3. Variation in the Base Composition of Transposable Elements

Early analyses of base composition in transposable elements were focused on the composition of the ORFs, in the context of codon bias. Multiple codons encode for the same amino acid, yet there is often a bias in the use of synonymous codons (i.e., codons that encode for the same amino acid), where some codons are preferred over others. This is a common phenomenon in eukaryotes that is referred to as "codon bias" and has been the subject of extensive attention by evolutionary biologists [64]. Codon bias may result from selection in favor of codons that are optimal in terms of translation accuracy [5,8] or efficiency, which is supported by the observation that strongly expressed genes exhibit a stronger codon bias than weakly expressed genes [6,7,65,66]. Because of this relation between codon bias and expression, it has been proposed that the analysis of codon bias in TEs could inform on the nature of the interactions between TEs and their hosts [14]. Early studies in *Drosophila* found that the codon usage of TEs differed from the codon usage of the host, with a bias in favor of codons ending in A or T [16]. Further studies in model organisms (*Arabidopsis thaliana*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Homo sapiens*) revealed a general AT-richness of TE's ORFs for both class I and class II elements compared with host genes [14] and a codon bias in TEs in favor of AT-ending codons, independently of the host. In most species, A-ending codons are preferred, except in the plants *A. thaliana* and *Oryza sativa*, where codons ending in T are preferred [17]. In general, the codon usage of TEs is different from the codon usage of host's genes but tends to be similar to that of weakly expressed genes, at least in some species [14,17]. These observations suggest two things. First, a general mechanism, common to all TEs and independent of the host, may be responsible for the AT-richness of TE's ORFs. Second, there is no tendency in TEs for codon optimization that would enhance the translation efficiency of the proteins they encode.

Although the general trend of an AT-richness and an AT-preference at the third position of codons seems to hold true for most TEs, more detailed analyses of a larger diversity of elements and the analysis of TEs in non-model organisms suggest a more nuanced and complex picture [15,67,68]. The analysis of base composition of non-LTR retrotransposons in vertebrates revealed large differences among clades of non-TR retrotransposons within the same genomes as well as large differences for the same clade among organisms [15]. For instance, in the lizard *Anolis carolinensis*, ORF2 (i.e., the ORF encoding for the reverse transcriptase) of elements belonging to the *L1* clade are enriched in AT (~67%) relative to host genes (~52% AT) while *L2* elements are GC-rich (~55% GC). In fish, *L1* and *L2* elements are AT-rich (~64% and ~58% AT, respectively) while elements of the *Rex1* clade (~52% GC) have a base composition close to the one of hosts genes (~50% AT). Although *L1* elements are universally AT-rich, there is a strong A bias on the positive strand in mammals and lizard (~41% and 43% A, respectively), a smaller bias in frogs (~33% A), and no bias in fish where A and T are equally represented. Interestingly, the base composition of the different clades is evolutionarily conserved and has persisted over long periods of evolutionary time within the same genome [15]. At the codon level, AT-rich codons are typically favored but there is no significant synonymous bias since the base frequency at the third position of codons fits the expectations given the overall nucleotide content of the sequences [15]. However, the codon usage of TEs tends to be closer to the codon usage of the host than expected given their base composition, which suggests a certain level of codon adaptation. Although TEs can be classified in AT-rich or GC-rich elements, some TEs show a highly unusual base composition. Such is the case of *L2* in the frog *Xenopus tropicalis* which is enriched in C (34%) and T (30%) on the positive strand [15]. These observations are not limited to vertebrates and similarly large differences in the GC% among class I elements were detected in the insect *Anopheles gambiae* [69].

Variation in base composition is not limited to class I elements. In a recent survey of GC content of TEs in fish, large variation in the base composition of class II elements was reported [68]. For example, class II elements in zebrafish are 36.9% GC while they are 44.1% GC in the pufferfish *Takifugu rubripes*. Interestingly, this study identified a positive

correlation between the genomic GC content and the TE GC content suggesting an effect of the overall genomic environment on the base composition of TEs.

In some unicellular organisms, the pattern of base composition is drastically different. For instance, in the choanoflagellate *Salpingoeca rosetta* [70], all TEs exhibit a preference for GC-ending codons and for translationally optimal codons, thus suggesting selection for translational efficiency. Similarly, in the stramenopile genus *Phytophthora* [71], LTR retrotransposons show preference for GC-ending codons that mirrors host genes. Although additional analyses of unicellular eukaryotes will be necessary, this observation suggests some differences between unicellular and multicellular organisms, perhaps related to different effective population size between these categories [60,70].

Different regions of TEs can also differ considerably in base composition and even ORFs from the same elements can exhibit different base composition. This is exemplified in the mammalian *L1* retrotransposons which have 5′UTRs (57.2 GC%) and 3′UTRs (46.3% GC) that are richer in GC than the two ORFs (39.1% for ORF1 and 37.9% for ORF2) [13,15]. The GC-richness of the L1 promoter is consistent with the nucleotide content found around transcription initiation sites in vertebrates, which is in part due to the abundance of CpG dinucleotides [72]. The AT-richness of ORF1 in vertebrate *L1* is always higher than ORF2, which is consistent with the fact that much more ORF1 protein is produced than ORF2 protein [73,74]. The difference between ORFs is even more striking for elements of the *L2* clade. In lizard, *L2* elements have a GC rich ORF2 (55% GC) but an AT rich ORF1 (54% AT) [15].

Finally, the base composition of non-autonomous elements is extremely variable and is not necessarily related to the base composition of the autonomous elements responsible for their mobility. This is exemplified for *SINE* elements that are mobilized by *LINE* elements. The *Alu* element in primates, which is mobilized by the AT-rich *L1* element, is GC-rich (63.3% GC for *AluY*) while the *SINE* elements in mouse are either AT-rich (e.g., *B2* elements; 52.2% AT) or GC-rich (e.g., *B1* elements; 59.9% GC).

4. Consequences of the Unusual Base Composition of Transposable Elements

The unusual base composition of TEs has a number of consequences for the mobility of TEs but also for the genome of their hosts. First the AT-richness of TEs, which is prevalent in multicellular organisms, is suboptimal for the transposition process, both at the transcriptional and at the translational level. In mammalian *L1* elements, the A-richness of the positive strand results in poor transcription because A-rich L1 sequences constitute a poor substrate for transcription elongation (either because of a slower rate of elongation, stalling of the RNA polymerase complex or premature dissociation) and because of the presence of A-rich premature poly-adenylation signals that are causing early transcription termination [4,75,76]. It should be noted however that the number of canonical poly-adenylation signals differs among non-LTR retrotransposons and is not directly related to the AT-richness since the number of predicted poly-adenylation signals could vary more than two folds among elements with the same base composition [15]. Although experimental data are lacking for most TEs, it is likely that the AT richness of most TEs impedes their efficient transcription, but the prediction that elements devoid of AT bias exhibit a more efficient transcription remains to be tested. The second potentially negative consequence of a high AT content is at the translational level. The prevalence of AT-ending codons in most TEs makes codon usage of their ORFs poorly adapted for efficient translation, which is supported by the similarity between the codon usage of TEs and weakly expressed host genes [14]. From the point of view of TEs, their AT-richness may appear maladaptive since it negatively affects their transcription and the translation of their proteins.

The negative effect of the biased base composition of TEs is not limited to the TEs but can also impact the expression of host genes in a number of ways. For instance, an AT-rich element inserted within a host gene could decrease the transcription of the gene either by reducing the efficiency of transcription or by producing prematurely terminated

transcripts [77,78]. This is one of the reasons AT-rich elements are rarely found in introns, and when they are, they tend to be oriented in the direction that is the least negative to gene expression [79–81]. This is exemplified in mammals where *L1* elements, which are AT-rich and have a strong A-bias on the positive strand, are extremely rare in introns, and the ones that have reached fixation are found in the opposite orientation to the host gene [79]. Another means TEs will affect the expression of host's genes is via epigenetic regulation [82]. Repression by DNA methylation at CpG sites constitutes the main means of defense against transposon activity in many organisms [49,50]. Although AT-rich elements will by definition contain few CpG sites, elements that are enriched in GC can contain a number of CpG sites that will be the target of methylation. The repressive mark can spread to the flanking sequences of the transposons and occasionally affect the expression of neighboring genes. The fact that methylated TEs are on average found further away from genes than unmethylated TEs [83,84] and tend to be at lower frequency in populations [85,86] is consistent with a negative effect of TE repression on their neighboring genes.

The base composition of TEs will also affect the overall genomic composition as well as the structure and function of the genome. It is well known that the abundance of TEs is the main determinant of the haploid genome size and TE amplification can cause rapid genome expansion [87–89], yet the impact of TEs on the base composition of the host has been underappreciated. In a recent study in fish, a positive correlation between the genomic base composition and the base composition of TEs was found [68]. Since most TEs are AT-rich, small genomes that contain few TEs tend to have a higher GC content than genomes that have experienced large TE amplifications. This observation suggests that the GC content of TEs will drive the GC content of the genomes in which they amplify. This observation is not limited to fish, and the amplification of AT-rich TEs in fungi can cause rapid changes in the genomic base composition. This is exemplified in the genus *Leptosphaeria* where strains that have experienced TE amplification have genomes with a lower GC content (45% GC) than strains that have not (51% GC) [90].

The GC content can differ among genomic regions (e.g., in birds, mammals and gars) or can be relatively homogenous (e.g., reptiles, amphibians and teleost fish) [1,2,91,92]. The cause of GC heterogeneity in birds and mammals has been the subject of extensive research. It is believed that the main driver of base content heterogeneity is a process called GC-biased gene conversion [93], which causes a fixation bias of G:C alleles over A:T alleles by a recombination-dependent process. Thus, regions of high recombination tend to be GC-rich while regions of low recombination tend to be GC-poor. An aspect that has received little attention is the contribution of TEs to GC heterogeneity. In mammals, the differential accumulation of TEs that differ in base composition contributes to the GC heterogeneity of the genome. AT-rich *L1* retrotransposons accumulate in regions of low recombination, presumably because they are eliminated by purifying selection from high recombining regions due to their ability to mediate ectopic recombination [54,94]. They will thus contribute to the higher AT content of regions of low recombination. In contrast, their non-autonomous counterpart, the *Alu* SINE, is GC-rich and tends to accumulate in genic regions with a high recombination rate [79], thus contributing to the evolution of these GC-rich genomic compartments. Although TEs certainly have an effect on regional base composition, this effect remains to be quantified and the recent development of tools that jointly analyze base composition and TE distribution will contribute to solving this gap in our knowledge [95].

5. Why Do Transposable Elements Have Such Unusual Base Composition?

Two main questions emerge from the analyses of the base composition of TEs. First, why do some TEs exhibit a nucleotide content that is so different from the genome average or from host genes? And second, why do some TEs from the same genome have drastically different composition? The unusual base composition of TEs could result from a number of non-exclusive factors that fall into three broad categories: horizontal transfer, mutational bias, and selective pressure.

TEs that have recently invaded a genome by horizontal transfer will exhibit a base composition that does not reflect processes which have taken place within the genome they occupy. In this case, we do not expect those TEs to show evidence of adaptation to the base composition of their new host. For this reason, a bias in the codon usage of TEs (which is related to the base composition) has been used as evidence of horizontal transfer [96–99], although many vertically transmitted TEs exhibit a similar bias [15,100,101]. Horizontal transfer can also explain differences in base composition between elements within the same genome. For instance, the genome of the medaka fish *Oryzias latipes* contains three families of *RTE* retrotransposons that differ substantially in base composition, but since *RTE* is prone to horizontal transfer [102,103], it is likely that these differences are caused by the independent transfer of *RTE*s from different sources [15].

The previous explanation does not apply to elements that are strictly (or mostly) vertically inherited. Many elements have persisted in genomes for very long periods of evolutionary time and have thus had time to evolve within the context of their host, and in this case, the evolution of their nucleotide content can be affected by mutational bias and/or selective pressure. Analyses of the pattern of mutation of recent copies of non-LTR retrotransposons, which are in majority AT-rich, revealed that mutations from C to T and G to A are the most abundant ones and that this mutational bias affects all elements, independently of their clade or base composition [15]. Although the overall mutational bias towards A and T is consistent with the general AT-richness of TEs, it fails to explain the strand bias observed for some elements such as *L1*. The cause of this mutational bias remains unclear but could result from a number of processes. The use of an error-prone reverse transcriptase by class I elements is unlikely to play a major role because the misincorporation of dATP by reverse transcriptase is exceedingly rare ([104], although this has only been tested on retroviral reverse transcriptase) and because some class I elements, like *L2*, use a reverse transcriptase for their transposition, yet they can have high GC content. In addition, this does not explain the high AT% of class II elements, which do not rely on a reverse transcriptase for their transposition. Another possibility comes from the action of DNA editing enzymes of the APOBEC family which are part of the defense system against viruses and retrotransposons. APOBEC3 proteins cause G to A mutations on the positive strand and could thus contribute to the A richness of some *LINE*s [105], although the signature of such editing was detected in a very small fraction of *L1* elements in humans. Interestingly, it was shown that APOBEC affects differently the AT-rich *L1* and the GC-rich *L2* elements in *Anolis carolinensis*; *L1* exhibited a signature of APOBEC editing while *L2* did not show any [106], a pattern consistent with the different base composition of these two clades of *LINE*s. More research is needed to assess the effect of editing on a broader range of organisms and to quantify the impact of APOBEC enzymes on base composition in a variety of contexts. Finally, the genomic environment in which elements are inserted could affect the type of mutations they experience. GC-biased gene conversion will affect differently elements inserted in regions with different recombination rate. Elements that accumulate in low recombining regions would be less subject to GC-biased gene conversion than elements that reside in regions of high recombination and will thus diverge in terms of nucleotide content. This process could be exacerbated by TE-specific insertion bias in favor of genomic regions with high or low recombination rate [107–109]. This hypothesis could be tested by comparing the mutation spectrum of TEs residing in different genomic compartments. It should be noted that the bias towards AT is not general, and for instance, a mutation bias from AT to GC was detected in the choanoflagelate *Salpingoeca rosetta* [70]. Interestingly, TEs in this species do not exhibit the AT-richness found in other organisms.

Another observation suggestive of a mutation-driven evolution of base composition comes from the correlation between the GC% of TEs and the GC% of non-repeated genomic DNA in fish [68]. This positive correlation suggests that the genomic context could be having an effect on the base composition of TEs, but it is unclear how the genomic GC% drives the GC% of TEs. A possibility is that, for their replication, TEs have to use the

pool of nucleotides available in the genome of their host, which thus constrains the base composition of TEs. Consequently, it is plausible that the composition of the pool of nucleotides will drive the evolution of the base composition of the TEs closer to the genomic average. This hypothesis remains to be tested, and the variation in base composition reported in teleost fish suggests that this group constitutes a good model.

A strictly neutralist mutational process is however unlikely to fully account for the unusual base composition of TEs, and a number of observations suggest that selection acts on base composition. First, in vertically transmitted TEs, such as *LINE*s, the base composition remains constant over long periods of evolutionary time suggesting selective pressure or functional constraint on elements [13,15]. This is exemplified in mammals where the *L1* retrotransposon has maintained its AT-richness and an A bias on the positive strand since the origin of this vertebrate class. Second, elements can differ in base composition within the same genome (e.g., *L1* and *L2* elements in *Anolis carolinensis*), although they are experiencing a similar pattern of mutation [15]. Third, different regions of the same element can exhibit a drastically different base composition. This is exemplified by the difference in base composition between the first and second ORFs of *L1* and *L2* in mammals and lizard [13,15]. Fourth, some TEs have retained a GC-rich or CT-rich content despite a mutational pressure towards A and T [15]. Fifth, in some species the codon usage is more similar to the host than expected given the base composition of elements, indicating a certain level of codon adaptation.

It is relatively easy to explain the base composition of TEs that exhibit GC% similar to their host's genes and those cases exemplify adaptation of TEs to their host [70]. It is much harder to provide a selectionist explanation for the apparently maladaptive base composition reported in many TEs. The first possibility is that the unusual nucleotide composition of TEs is not detrimental but is in fact adaptive and is responsible for the fine-tuning of TEs' transcription and translation. In this context, it is plausible that the high AT content is a mechanism of self-regulation reflecting a trade-off between the efficiency of transposition and the negative impact of transposition on the host [75]. For instance, selection could favor an inefficient transcription of TEs because an element that would be transcribed too efficiently could transpose at a level that would be detrimental to the host [76]. Selection at the transcriptional level seems more likely than selection at the translational level since the AT-richness is often observed at the three codon positions [15]. It is not to say that selection acts only to reduce the efficacy of transposition. In some species, the codon usage tends to be more similar to the host than expected given the composition of the element, which is evidence in favor of adaptation to the translational context of the host genome. This could also explain why the base composition may differ among regions of TEs. For instance, the ORF1 of *LINE*s has usually a less biased base composition than ORF2, possibly because ORF1p needs to be produced in a much larger amount than ORF2p for successful transposition [73,74]. Another source of selection is that the high AT content is a means of escaping repression by the host. TE expression is regulated by methylation of cytosine at CpG, and a low GC content could constitute a defense of TEs against such inactivating mechanisms. The general under-representation of the CpG dinucleotide in TEs is consistent with this possibility [15,110].

6. Conclusions

Understanding the causes of the unusual base composition of TEs and the respective role of mutational bias and selective pressure remains an understudied aspect of TE biology. It can however teach us a lot about the nature of the interactions between TEs and their hosts. The long-term persistence of a suboptimal base composition could support a model of coexistence between TEs and host, whereby elements evolve towards a reduced transposition rate that minimizes the negative impact they can have on their host. It has been proposed that such model may be more prevalent than the arms-race model that has prevailed until recently [111], yet additional studies will be necessary to confirm that the biased composition of TEs is in fact adaptive. A second unexplored aspect of TE biology

that could be related to base composition is how different TEs coexist and possibly compete in their genomic environment. By analogy with the field of ecology, it has been proposed that TEs are comparable with organisms that are sharing a genomic habitat within which they interact [112,113]. The coexistence within the same genomes of TEs with different base composition could inform on the nature of these interactions. For instance, it is possible that elements with different GC content do not use the same resources (either tRNA or amino acids) and could thus coexist since they are not using the same "genomic niche".

A better understanding of the compositional bias will require further studies. In particular, comparing the timing and intensity of expression of TEs that differ in base composition, within the same genome, could potentially inform on the functionality and possible adaptive value of a biased base composition. Experimental approaches that synthetically modify the base composition of elements by optimizing or de-optimizing codon usage could also prove informative. Finally, we should not underestimate how biased in favor of model organisms our knowledge of TEs is. Recent investigations on unicellular organisms [70] for instance have challenged the common belief that all TEs are AT rich, and we can expect that studies on more non-model organisms will also bring their share of surprises.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Costantini, M.; Cammarano, R.; Bernardi, G. The evolution of isochore patterns in vertebrate genomes. *BMC Genom.* **2009**, *10*, 146. [CrossRef]
2. Bernardi, G.; Bernardi, G. Compositional patterns in the nuclear genome of cold-blooded vertebrates. *J. Mol. Evol.* **1990**, *31*, 265–281. [CrossRef]
3. Zhou, Z.; Dang, Y.; Zhou, M.; Li, L.; Yu, C.-H.; Fu, J.; Chen, S.; Liu, Y. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E6117–E6125. [CrossRef]
4. Han, J.S.; Szak, S.T.; Boeke, J.D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **2004**, *429*, 268–274. [CrossRef] [PubMed]
5. Akashi, H. Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy. *Genetics* **1994**, *136*, 927–935. [CrossRef] [PubMed]
6. Duret, L. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* **2002**, *12*, 640–649. [CrossRef]
7. Duret, L.; Mouchiroud, D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4482–4487. [CrossRef]
8. Stoletzki, N.; Eyre-Walker, A. Synonymous codon usage in *Escherichia coli*: Selection for translational accuracy. *Mol. Biol. Evol.* **2007**, *24*, 374–381. [CrossRef] [PubMed]
9. Almojil, D.; Bourgeois, Y.; Falis, M.; Hariyani, I.; Wilcox, J.; Boissinot, S. The structural, functional and evolutionary impact of transposable elements in eukaryotes. *Genes* **2021**, *12*, 918. [CrossRef]
10. Capy, P. Taming, Domestication and Exaptation: Trajectories of Transposable Elements in Genomes. *Cells* **2021**, *10*, 3590. [CrossRef]
11. Bourque, G.; Burns, K.H.; Gehring, M.; Gorbunova, V.; Seluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H.L.; Macfarlan, T.S. Ten things you should know about transposable elements. *Genome Biol.* **2018**, *19*, 199. [CrossRef] [PubMed]
12. Sinzelle, L.; Izsvak, Z.; Ivics, Z. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. *Cell. Mol. Life Sci.* **2009**, *66*, 1073–1093. [CrossRef] [PubMed]
13. Boissinot, S.; Sookdeo, A. The evolution of LINE-1 in vertebrates. *Genome Biol. Evol.* **2016**, *8*, 3485–3507. [CrossRef]
14. Lerat, E.; Capy, P.; Biemont, C. Codon usage by transposable elements and their host genes in five species. *J. Mol. Evol.* **2002**, *54*, 625–637. [CrossRef]
15. Ruggiero, R.P.; Boissinot, S. Variation in base composition underlies functional and evolutionary divergence in non-LTR retrotransposons. *Mob. DNA* **2020**, *11*, 14–18. [CrossRef]

16. Shields, D.C.; Sharp, P.M. Evidence that mutation patterns vary among *Drosophila* transposable elements. *J. Mol. Biol.* **1989**, *207*, 843–846. [CrossRef]

17. Jia, J.; Xue, Q. Codon usage biases of transposable elements and host nuclear genes in *Arabidopsis thaliana* and *Oryza sativa*. *Genom. Proteom. Bioinform.* **2009**, *7*, 175–184. [CrossRef]

18. Andrieu, O.; Fiston, A.-S.; Anxolabéhère, D.; Quesneville, H. Detection of transposable elements by their compositional bias. *BMC Bioinform.* **2004**, *5*, 94. [CrossRef]

19. Tollis, M.; Boissinot, S. The evolutionary dynamics of transposable elements in eukaryote genomes. *Repetitive DNA* **2012**, *7*, 68–91.

20. Wells, J.N.; Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* **2020**, *54*, 539–561. [CrossRef]

21. Kojima, K.K. Structural and sequence diversity of eukaryotic transposable elements. *Genes Genet. Syst.* **2018**, *94*, 233–252. [CrossRef] [PubMed]

22. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [CrossRef] [PubMed]

23. Eickbush, T.H.; Malik, H.S. Origins and evolution of retrotransposons. In *Mobile DNA II*; American Society of Microbiology: Washington, DC, USA, 2002; pp. 1111–1144.

24. Dewannieux, M.; Esnault, C.; Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **2003**, *35*, 41–48. [CrossRef] [PubMed]

25. Dewannieux, M.; Heidmann, T. LINEs, SINEs and processed pseudogenes: Parasitic strategies for genome modeling. *Cytogenet. Genome Res.* **2005**, *110*, 35–48. [CrossRef]

26. Ohshima, K.; Okada, N. SINEs and LINEs: Symbionts of eukaryotic genomes with a common tail. *Cytogenet. Genome Res.* **2005**, *110*, 475–490. [CrossRef]

27. Feschotte, C.; Pritham, E.J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **2007**, *41*, 331–368. [CrossRef]

28. Kapitonov, V.V.; Jurka, J. Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8714–8719. [CrossRef]

29. Kapitonov, V.V.; Jurka, J. Helitrons on a roll: Eukaryotic rolling-circle transposons. *Trends Genet.* **2007**, *23*, 521–529. [CrossRef]

30. Kapitonov, V.V.; Jurka, J. Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4540–4545. [CrossRef]

31. Pritham, E.J.; Putliwala, T.; Feschotte, C. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **2007**, *390*, 3–17. [CrossRef]

32. Hartl, D.; Lozovskaya, E.; Lawrence, J. Nonautonomous transposable elements in prokaryotes and eukaryotes. *Genetica* **1992**, *86*, 47–53. [CrossRef] [PubMed]

33. Novick, P.A.; Smith, J.D.; Floumanhaft, M.; Ray, D.A.; Boissinot, S. The evolution and diversity of DNA transposons in the genome of the lizard *Anolis carolinensis*. *Genome Biol. Evol.* **2011**, *3*, 1–14. [CrossRef] [PubMed]

34. Yang, G.; Nagel, D.H.; Feschotte, C.; Hancock, C.N.; Wessler, S.R. Tuned for transposition: Molecular determinants underlying the hyperactivity of a Stowaway MITE. *Science* **2009**, *325*, 1391–1394. [CrossRef] [PubMed]

35. Smit, A.F.; Tóth, G.; Riggs, A.D.; Jurka, J. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **1995**, *246*, 401–417. [CrossRef] [PubMed]

36. Furano, A.V. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Prog. Nucleic Acid Res. Mole. Biol.* **2000**, *64*, 255–294.

37. Furano, A.V.; Duvernell, D.D.; Boissinot, S. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* **2004**, *20*, 9–14. [CrossRef]

38. Khan, H.; Smit, A.; Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **2006**, *16*, 78–87. [CrossRef]

39. Sookdeo, A.; Hepp, C.M.; Boissinot, S. Contrasted patterns of evolution of the LINE-1 retrotransposon in perissodactyls: The history of a LINE-1 extinction. *Mob. DNA* **2018**, *9*, 12. [CrossRef]

40. Sookdeo, A.; Hepp, C.M.; McClure, M.A.; Boissinot, S. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob. DNA* **2013**, *4*, 3. [CrossRef]

41. Boissinot, S.; Furano, A.V. Adaptive evolution in LINE-1 retrotransposons. *Mol. Biol. Evol.* **2001**, *18*, 2186–2194. [CrossRef]

42. Jacobs, F.M.; Greenberg, D.; Nguyen, N.; Haeussler, M.; Ewing, A.D.; Katzman, S.; Paten, B.; Salama, S.R.; Haussler, D. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **2014**, *516*, 242–245. [CrossRef] [PubMed]

43. Blumenstiel, J.P. Birth, school, work, death, and resurrection: The life stages and dynamics of transposable element proliferation. *Genes* **2019**, *10*, 336. [CrossRef] [PubMed]

44. Schaack, S.; Gilbert, C.; Feschotte, C. Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **2010**, *25*, 537–546. [CrossRef] [PubMed]

45. Gilbert, C.; Feschotte, C. Horizontal acquisition of transposable elements and viral sequences: Patterns and consequences. *Curr. Opin. Genet. Dev.* **2018**, *49*, 15–24. [CrossRef]

46. Bourgeois, Y.; Boissinot, S. On the population dynamics of junk: A review on the population genomics of transposable elements. *Genes* **2019**, *10*, 419. [CrossRef]

47. Goodier, J.L. Restricting retrotransposons: A review. *Mob. DNA* **2016**, *7*, 16. [CrossRef]

48. Czech, B.; Hannon, G.J. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem. Sci.* **2016**, *41*, 324–337. [CrossRef]

49. Deniz, O.; Frost, J.M.; Branco, M.R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **2019**, *20*, 417–431. [CrossRef]

50. Yoder, J.A.; Walsh, C.P.; Bestor, T.H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **1997**, *13*, 335–340. [CrossRef]

51. Bourgeois, Y.; Ruggiero, R.; Hariyani, I.; Boissinot, S. Disentangling the determinants of transposable elements dynamics in vertebrate genomes using empirical evidences and simulations. *PLoS Genet.* **2020**, *16*, e1009082. [CrossRef]

52. Ruggiero, R.P.; Bourgeois, Y.; Boissinot, S. LINE insertion polymorphisms are abundant but at low frequencies across populations of *Anolis carolinensis*. *Front. Genet.* **2017**, *8*, 44. [CrossRef] [PubMed]

53. Biémont, C.; Lemeunier, F.; Guerreiro, M.G.; Brookfield, J.; Gautier, C.; Aulard, S.; Pasyukova, E. Population dynamics of the copia, mdg1, mdg3, gypsy, and P transposable elements in a natural population of *Drosophila melanogaster*. *Genet. Res.* **1994**, *63*, 197–212. [CrossRef] [PubMed]

54. Boissinot, S.; Entezam, A.; Furano, A.V. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* **2001**, *18*, 926–935. [CrossRef]

55. Charlesworth, B.; Langley, C.H. The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* **1989**, *23*, 251–287. [CrossRef] [PubMed]

56. Xue, A.T.; Ruggiero, R.P.; Hickerson, M.J.; Boissinot, S. Differential effect of selection against LINE retrotransposons among vertebrates inferred from whole-genome data and demographic modeling. *Genome Biol. Evol.* **2018**, *10*, 1265–1281. [CrossRef] [PubMed]

57. Gonzalez, J.; Macpherson, J.M.; Messer, P.W.; Petrov, D.A. Inferring the strength of selection in *Drosophila* under complex demographic models. *Mol. Biol. Evol.* **2009**, *26*, 513–526. [CrossRef] [PubMed]

58. Lockton, S.; Ross-Ibarra, J.; Gaut, B.S. Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13965–13970. [CrossRef] [PubMed]

59. Garcia Guerreiro, M.P.; Chavez-Sandoval, B.E.; Balanya, J.; Serra, L.; Fontdevila, A. Distribution of the transposable elements bilbo and gypsy in original and colonizing populations of *Drosophila subobscura*. *BMC Evol. Biol.* **2008**, *8*, 234. [CrossRef]

60. Lynch, M.; Conery, J.S. The origins of genome complexity. *Science* **2003**, *302*, 1401–1404. [CrossRef]

61. Blass, E.; Bell, M.; Boissinot, S. Accumulation and rapid decay of non-LTR retrotransposons in the genome of the three-spine stickleback. *Genome Biol. Evol.* **2012**, *4*, 687–702. [CrossRef]

62. Petrov, D.A.; Sangster, T.A.; Johnston, J.S.; Hartl, D.L.; Shaw, K.L. Evidence for DNA loss as a determinant of genome size. *Science* **2000**, *287*, 1060–1062. [CrossRef] [PubMed]

63. Kapusta, A.; Suh, A.; Feschotte, C. Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E1460–E1469. [CrossRef] [PubMed]

64. Hershberg, R.; Petrov, D.A. Selection on codon bias. *Annu. Rev. Genet.* **2008**, *42*, 287–299. [CrossRef] [PubMed]

65. Ingvarsson, P.K. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Mol. Biol. Evol.* **2007**, *24*, 836–844. [CrossRef]

66. Shields, D.C.; Sharp, P.M.; Higgins, D.G.; Wright, F. "Silent" sites in Drosophila genes are not neutral: Evidence of selection among synonymous codons. *Mol. Biol. Evol.* **1988**, *5*, 704–716.

67. Gaffaroglu, M.; Majtanova, Z.; Symonova, R.; Pelikanova, S.; Unal, S.; Lajbner, Z.; Rab, P. Present and Future Salmonid Cytogenetics. *Genes* **2020**, *11*, 1462. [CrossRef]

68. Symonova, R.; Suh, A. Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes. *Mob. DNA* **2019**, *10*, 49. [CrossRef]

69. Besansky, N. Codon usage patterns in chromosomal and retrotransposon genes of the mosquito *Anopheles gambiae*. *Insect Mol. Biol.* **1993**, *1*, 171–178. [CrossRef]

70. Southworth, J.; Grace, C.A.; Marron, A.O.; Fatima, N.; Carr, M. A genomic survey of transposable elements in the choanoflagellate *Salpingoeca rosetta* reveals selection on codon usage. *Mob. DNA* **2019**, *10*, 44. [CrossRef]

71. Jiang, R.H.; Govers, F. Nonneutral GC3 and retroelement codon mimicry in *Phytophthora*. *J. Mol. Evol.* **2006**, *63*, 458–472. [CrossRef]

72. Aerts, S.; Thijs, G.; Dabrowski, M.; Moreau, Y.; De Moor, B. Comprehensive analysis of the base composition around the transcription start site in Metazoa. *BMC Genom.* **2004**, *5*, 34. [CrossRef] [PubMed]

73. Basame, S.; Li, P.W.-l.; Howard, G.; Branciforte, D.; Keller, D.; Martin, S.L. Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition. *J. Mol. Biol.* **2006**, *357*, 351–357. [CrossRef]

74. Doucet, A.J.; Hulme, A.E.; Sahinovic, E.; Kulpa, D.A.; Moldovan, J.B.; Kopera, H.C.; Athanikar, J.N.; Hasnaoui, M.; Bucheton, A.; Moran, J.V. Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet.* **2010**, *6*, e1001150. [CrossRef] [PubMed]

75. Perepelitsa-Belancio, V.; Deininger, P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.* **2003**, *35*, 363–366. [CrossRef] [PubMed]

76. Han, J.S.; Boeke, J.D. A highly active synthetic mammalian retrotransposon. *Nature* **2004**, *429*, 314–318. [CrossRef] [PubMed]

77. Han, J.S.; Boeke, J.D. LINE-1 retrotransposons: Modulators of quantity and quality of mammalian gene expression? *Bioessays* **2005**, *27*, 775–784. [CrossRef] [PubMed]

78. Roy-Engel, A.; El-Sawy, M.; Farooq, L.; Odom, G.; Perepelitsa-Belancio, V.; Bruch, H.; Oyeniran, O.; Deininger, P. Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet. Genome Res.* **2005**, *110*, 365–371. [CrossRef]

79. Medstrand, P.; van de Lagemaat, L.N.; Mager, D.L. Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Res.* **2002**, *12*, 1483–1495. [CrossRef]

80. Cutter, A.D.; Good, J.M.; Pappas, C.T.; Saunders, M.A.; Starrett, D.M.; Wheeler, T.J. Transposable element orientation bias in the *Drosophila melanogaster* genome. *J. Mol. Evol.* **2005**, *61*, 733–741. [CrossRef]

81. van de Lagemaat, L.N.; Landry, J.R.; Mager, D.L.; Medstrand, P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* **2003**, *19*, 530–536. [CrossRef]

82. Slotkin, R.K.; Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **2007**, *8*, 272–285. [CrossRef] [PubMed]

83. Kinoshita, Y.; Saze, H.; Kinoshita, T.; Miura, A.; Soppe, W.J.; Koornneef, M.; Kakutani, T. Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J.* **2007**, *49*, 38–45. [CrossRef] [PubMed]

84. Parinov, S.; Sundaresan, V. Functional genomics in *Arabidopsis*: Large-scale insertional mutagenesis complements the genome sequencing project. *Curr. Opin. Biotechnol.* **2000**, *11*, 157–161. [CrossRef]

85. Hollister, J.D.; Gaut, B.S. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **2009**, *19*, 1419–1428. [CrossRef] [PubMed]

86. Lee, Y.C.G.; Karpen, G.H. Pervasive epigenetic effects of *Drosophila* euchromatic transposable elements impact their evolution. *eLife* **2017**, *6*, e25762. [CrossRef]

87. Vitte, C.; Panaud, O. LTR retrotransposons and flowering plant genome size: Emergence of the increase/decrease model. *Cytogenet. Genome Res.* **2005**, *110*, 91–107. [CrossRef]

88. Hawkins, J.S.; Kim, H.; Nason, J.D.; Wing, R.A.; Wendel, J.F. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* **2006**, *16*, 1252–1261. [CrossRef]

89. Piegu, B.; Guyot, R.; Picault, N.; Roulin, A.; Sanyal, A.; Kim, H.; Collura, K.; Brar, D.S.; Jackson, S.; Wing, R.A.; et al. Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* **2006**, *16*, 1262–1269. [CrossRef]

90. Grandaubert, J.; Lowe, R.G.; Soyer, J.L.; Schoch, C.L.; Van de Wouw, A.P.; Fudal, I.; Robbertse, B.; Lapalu, N.; Links, M.G.; Ollivier, B. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC Genom.* **2014**, *15*, 891. [CrossRef]

91. Symonova, R.; Majtanova, Z.; Arias-Rodriguez, L.; Morkovsky, L.; Korinkova, T.; Cavin, L.; Pokorna, M.J.; Dolezalkova, M.; Flajshans, M.; Normandeau, E.; et al. Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish. *J. Exp. Zool. B Mol. Dev. Evol.* **2017**, *328*, 607–619. [CrossRef]

92. Costantini, M.; Auletta, F.; Bernardi, G. Isochore patterns and gene distributions in fish genomes. *Genomics* **2007**, *90*, 364–371. [CrossRef] [PubMed]

93. Mugal, C.F.; Weber, C.C.; Ellegren, H. GC-biased gene conversion links the recombination landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. *Bioessays* **2015**, *37*, 1317–1326. [CrossRef] [PubMed]

94. Song, M.; Boissinot, S. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene* **2007**, *390*, 206–213. [CrossRef] [PubMed]

95. Matoulek, D.; Borůvková, V.; Ocalewicz, K.; Symonová, R. GC and repeats profiling along chromosomes—The future of fish compositional cytogenomics. *Genes* **2021**, *12*, 50. [CrossRef] [PubMed]

96. Rödelsperger, C.; Sommer, R.J. Computational archaeology of the *Pristionchus pacificus* genome reveals evidence of horizontal gene transfers from insects. *BMC Evol. Biol.* **2011**, *11*, 239. [CrossRef]

97. McHale, M.T.; Roberts, I.N.; Noble, S.M.; Beaumont, C.; Whitehead, M.P.; Seth, D.; Oliver, R.P. CfT-I: An LTR-retrotransposon in Cladosporium fulvum, a fungal pathogen of tomato. *Mol. Gen. Genet. MGG* **1992**, *233*, 337–347. [CrossRef]

98. Wallau, G.L.; Ortiz, M.F.; Loreto, E.L.S. Horizontal transposon transfer in eukarya: Detection, bias, and perspectives. *Genome Biol. Evol.* **2012**, *4*, 801–811. [CrossRef]

99. Powell, J.R.; Gleason, J.M. Codon usage and the origin of P elements. *Mol. Biol. Evol.* **1996**, *13*, 278–279. [CrossRef]

100. Springer, M.S.; Tusneem, N.A.; Davidson, E.H.; Britten, R.J. Phylogeny, rates of evolution, and patterns of codon usage among sea urchin retroviral-like elements, with implications for the recognition of horizontal transfer. *Mol. Biol. Evol.* **1995**, *12*, 219–230.

101. Lerat, E.; Biémont, C.; Capy, P. Codon usage and the origin of P elements. *Mol. Biol. Evol.* **2000**, *17*, 467–468. [CrossRef]

102. Kordis, D.; Gubensek, F. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10704–10709. [CrossRef] [PubMed]

103. Walsh, A.M.; Kortschak, R.D.; Gardner, M.G.; Bertozzi, T.; Adelson, D.L. Widespread horizontal transfer of retrotransposons. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1012–1016. [CrossRef]

104. Preston, B.D.; Poiesz, B.J.; Loeb, L.A. Fidelity of HIV-1 reverse transcriptase. *Science* **1988**, *242*, 1168–1171. [CrossRef] [PubMed]

105. Carmi, S.; Church, G.M.; Levanon, E.Y. Large-scale DNA editing of retrotransposons accelerates mammalian genome evolution. *Nat. Commun.* **2011**, *2*, 519. [CrossRef] [PubMed]

106. Lindič, N.; Budič, M.; Petan, T.; Knisbacher, B.A.; Levanon, E.Y.; Lovšin, N. Differential inhibition of LINE1 and LINE2 retrotransposition by vertebrate AID/APOBEC proteins. *Retrovirology* **2013**, *10*, 156. [CrossRef]

107. Duret, L.; Marais, G.; Biémont, C. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in Caenorhabditis elegans. *Genetics* **2000**, *156*, 1661–1669. [CrossRef]

108. Kawakami, T.; Mugal, C.F.; Suh, A.; Nater, A.; Burri, R.; Smeds, L.; Ellegren, H. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Mol. Ecol.* **2017**, *26*, 4158–4172. [CrossRef]

109. Myers, S.; Freeman, C.; Auton, A.; Donnelly, P.; McVean, G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.* **2008**, *40*, 1124. [CrossRef]

110. Lerat, E.; Capy, P.; Biémont, C. The relative abundance of dinucleotides in transposable elements in five species. *Mol. Biol. Evol.* **2002**, *19*, 964–967. [CrossRef]

111. Cosby, R.L.; Chang, N.-C.; Feschotte, C. Host–transposon interactions: Conflict, cooperation, and cooption. *Genes Dev.* **2019**, *33*, 1098–1116. [CrossRef]

112. Venner, S.; Feschotte, C.; Biémont, C. Dynamics of transposable elements: Towards a community ecology of the genome. *Trends Genet.* **2009**, *25*, 317–323. [CrossRef] [PubMed]

113. Brookfield, J.F. The ecology of the genome—Mobile DNA elements and their hosts. *Nat. Rev. Genet.* **2005**, *6*, 128–136. [CrossRef] [PubMed]