Impact of transposable elements on the evolution of complex living systems and their epigenetic control

Ambra Viviani, Maria Ventimiglia, Marco Fambrini, Alberto Vangelisti, Flavia Mascagni, Claudio Pugliesi, Gabriele Usai

Department of Agriculture, Food and Environment (DAFE), University of Pisa, Via del Borghetto, 80-56124 Pisa, Italy

Ambra Viviani and Maria Ventimiglia contributed equally to this work.

Corresponding Author: Claudio Pugliesi, Email: claudio.pugliesi@unipi.it

Abstract

Transposable elements (TEs) contribute to genomic innovations, as well as genome instability, across a wide variety of species. Popular designations such as ‘selfish DNA’ and ‘junk DNA,’ common in the 1980s, may be either inaccurate or misleading, while a more enlightened view of the TE-host relationship covers a range from parasitism to mutualism. Both plant and animal hosts have evolved epigenetic mechanisms to reduce the impact of TEs, both by directly silencing them and by reducing their ability to transpose in the genome. However, TEs have also been co-opted by both plant and animal genomes to perform a variety of physiological functions, ranging from TE-derived proteins acting directly in normal biological functions to innovations in transcription factor activity and also influencing gene expression. Their presence, in fact, can affect a range of features at genome, phenotype, and population levels. The impact that TEs have had on evolution is multifaceted, and many aspects still remain unexplored. In this review, the epigenetic control of TEs is contextualised according to the evolution of complex living systems.

Keywords

Analog information; Complex living systems; Digital information; Transposable elements; Epigenetic mechanism

1. Introduction

Transposable elements (TEs) are DNA sequences capable of changing position within the host genome (McClintock, 1984; Huang et al., 2012; Bourque et al., 2018; Koonin et al., 2020). Their repositioning occurs directly through a 'cut and paste' mechanism or indirectly through a 'copy and paste' mechanism which provides the formation of an RNA intermediate (retrotransposons, REs). By virtue of these mechanisms, TEs can disperse within single genomes, between genomes of the same population, or even between genomes of different species (Koonin et al., 2020; Gill et al., 2021). In most prokaryotes and eukaryotes, TEs constitute the preponderant fraction of the entire genome (Bourque et al., 2018; Neumann et al., 2019; Fambrini et al., 2020; Gill et al., 2021) (Fig. 1).

Many excellent reviews are available on the classification, mobilisation, and impact of TEs on both genetic and genomic variability (Feschotte, 2008; Pace et al., 2008; Lisch, 2013; Oliver et al., 2013; Ayarpadikannan and Kim, 2014; Chuong et al., 2017; Bourque et al., 2018; Dubin et al., 2018; Cosby et al., 2019; Neumann et al., 2019; Orozco-Arias et al., 2019; Fambrini et al., 2020; Koonin et al., 2020; Almojil et al., 2021; Gill et al., 2021). Nevertheless, the main focuses of this review are the discovery of TEs, the epigenetic control of these elements, and how these sequences have impacted evolution processes in a complex system, like a living organism (Fig.1).

To understand the role that TEs have played during evolution, it is essential to consider both the history of their discovery and their mechanisms of action. It is necessary to reconstruct the path that, over a few decades, demonstrated the presence of TEs in an increasing number of organisms (McClintock, 1950; Sotero-Caio et al., 2017; Dubin et al., 2018; Neumann et al., 2019; Orozco-Arias et al., 2019; Cain et al., 2020; Fambrini et al., 2020). Furthermore, it is important to understand how the progress of modern sequencing technologies has made it possible to explain the influence they have exerted on spontaneous mutations and on the regulation of gene expression during evolution (Warren et al., 2015; Bourque et al., 2018; Neumann et al., 2019; Orozco-Arias et al., 2019).

According to Kuhn (1962), scientific progress does not evolve by virtue of an accumulation of direct observations, but due both to the advancement of experimental technologies and to the progressive theoretical elaboration of their interpretations. The ultimate resolution of this process is the inclusion of available information in a single

paradigm/theory capable of unifying old and new data in a coherent logical framework. The objective is achieved when the initial contradictions expressed by alternative theories are resolved in the prevalence of a dominant paradigm, which sometimes requires a long time before its acceptance. Mendel's discoveries had to wait over thirty years before their importance was recognised and the implications analysed (Singh, 2015). The scientific work of Barbara McClintock (1902–1992), who discovered TEs (McClintock, 1950), is also exemplary in this sense because it summarises all the phases of a path in the history of science that leads from the initial discovery to the recognition and acceptance of the validity of her interpretations (McClintock, 1984).

It should also be considered that from the initial scepticism, the subsequent acceptance of the role of TEs in the genomic evolution of various organisms reached unimaginable evolutionary connotations (Fambrini et al., 2020). Furthermore, concepts such as symbiosis, conflict, cooperation, co-optation and exaptation between TEs and their host have been developed (Labrador and Corces, 1997; Joly-Lopez and Bureau, 2018; Cosby et al., 2019). Many examples are now emerging showing that TEs can contribute significantly to the well-being of the organisms they populate (Fig. 1 and 2). Nevertheless, in the host-TE interaction, epigenetic controls on their mobilisation, localisation, and inactivation have evolved and continue to evolve (Cosby et al., 2019; Fambrini et al., 2020).

2. Discovery of transposable elements

Native Americans have cultivated the corn plant (*Zea mays* subsp. *mays*) for thousands of years. When the Europeans arrived in America, they were impressed by the great variability of colours in corn ears and their kernels. In the 1930s and 1940s, the colour variegation of corn kernels attracted the attention of many researchers, among which the most active was certainly Barbara McClintock. In collaboration with and stimulated by Marcus Rhoades (Rhoades, 1936), McClintock began studying these phenotypes to understand their genetic basis. The two researchers were impressed by the widespread and uneven mosaicism of some characteristics, from the non-Mendelian distribution of segregating phenotypes in the hybrid progenies, and by the high rate of somatic and germinal reversion of the mutations. Many years of experimental work allowed McClintock to identify genetic elements as responsible for some of the observed

phenomena, among which *Dissociation* (*Ds*) and *Activator* (*Ac*) constituted the first and one of the most representative examples. In particular, she observed that whenever the *Ds* element was inserted in particular chromosomal regions, it produced a variety of phenotypic effects on the kernels, attributable to insertional mutations or chromosomal deletions. A further series of crossings and cytological observations allowed her to observe how the *Ac* element controlled the transposition of *Ds* and how this mobilisation could sometimes be associated with breakdown events of the chromosome itself (McClintock, 1950). It is in coincidence with these transposition events that alterations in the synthesis of starch and in the deposition of pigments were determined with the phenomena of mosaicism due to the colour and shape of the kernels. To interpret these observations, McClintock correctly hypothesised that the *Ds* and *Ac* sequences, which she defined as 'controlling elements,' could act as mobile chromosomal regions capable of modulating or even suppressing the expression of genes at different times in the development of the kernel (McClintock, 1950; 1953; 1956).

Subsequent molecular analyses have shown that *Ds* is a non-autonomous element derived from *Ac* due to the loss of some fundamental regions upon transposition (Fedoroff et al., 1983). The numerous cytogenetic analyses carried out in support of these experiments also led McClintock to discover other controlling elements involved in the mechanism of chromosomal instability (McClintock, 1953; 1956). The conformity of the genetic and cytological results eventually led her to develop a dynamic view of the genome, according to which entire gene portions can move from one chromosome to another, generating gene inactivation, gene deletions, or more extensive chromosomal mutations (Xiao et al., 2008; Yu et al., 2011; Cerbin and Jiang, 2018). Chromosomes are no longer rigid structures but show flexibility, thereby allowing reorganisation of the genetic material (Underwood et al., 2019).

For many decades, the scientific community has viewed McClintock's work with great scepticism. Not only was the role of TEs considered marginal, with respect to genetics known in the first half of the twentieth century, but it was also the same interpretation proposed by McClintock that was hindered because it was incompatible with the dominant paradigms in the 1950s and 1960s. The genetics of those years were mainly centred on the study of allozyme polymorphism to decipher the structure and phylogeny of a population. The research tended to identify any possible mutation that modified or altered the coding function of allozymes and that could therefore itself be considered a primary

source of evolutionary change (Biémont, 2010). Ultimately, it was the study of mutant alleles that allowed the genetic population approach to define the central paradigm of early twentieth century genetics (Rolls-Hansen, 2014). The invariance of gene recombination frequencies was considered a demonstration of the stability of the related genes in a chromosomal position. Following this interpretation, genes were conceived as static loci placed on the chromosome according to a linear succession as beads on a string. Further support for this chromosomal schematisation was provided by the hypothesis of Beadle and Tatum (1941), according to which each gene is correlated with the production of a specific enzyme. Both these concepts were required to justify the discovery of new genetic elements in functional terms, based on the criteria of presumed usefulness for the organism that was the bearer. McClintock also suggested a parallelism between the 'controlling elements' and the operon control systems in bacteria (Jacob and Monod, 1961) and maize (McClintock, 1961). In contrast, Jacob and Monod referred to 'controlling elements' as extrachromosomal episomes that were perhaps viral in origin but not as components of the normal regulatory equipment of cells (Comfort, 2001). It is probably for these reasons that, unable to attribute any role to the existence and diffusion of TEs, the scientific community tried to denigrate their importance. Therefore, for a long time, TEs were labelled selfish or junk DNA due to their alleged evolutionary neutrality or parasitic nature (Orgel and Crick, 1980). Conversely, TEs have played and will continue to play important evolutionary roles despite their often deleterious effects on gene activity (Fig. 1).

Many spontaneous mutations with important evolutionary features in plants [e.g., regulation of *TEOSINTE BRANCHED1* (*TB1*) gene during maize domestication (Studer et al., 2011) and animals (e.g., the origin of *Biston betularia* subsp. *carbonifera*) during the industrial revolution of the 18th century (Van't Hof et al., 2016) are explicative of the key role of TE activity. In maize, a RE (*Hopscotch*) inserted into an enhancer region of the *TB1* gene increased the gene expression of *TB1* (Studer et al., 2011). The insertion, approximately 60 kb from the *TB1* locus*,* determined the increase in apical dominance and a different position of male and female inflorescences, in respect to its wild ancestor teosinte (Clark et al., 2006). In *B. betularia,* the event that gave rise to industrial melanism was the insertion of a large (9 kb) TE (*carb*-TE), repeated in tandem, into the first intron of the *CORTEX* gene (Van'tHof et al., 2016). Outstandingly, the wrinkled-seed trait of pea described by Mendel in his studies on the inheritance of characters is caused by a TE-like (i.e., a non-autonomous TE element) insertion in a gene encoding starch-branching enzyme isoform 1 (SBE1) (Bhattacharyya et al., 1990; 1993). Many other examples have

been described (Volff, 2006: Bourque et al., 2018; Dubin et al., 2018; Neumann et al., 2019; Orozco-Arias et al., 2019; Fambrini et al., 2020; Almojil et al., 2021). Therefore, it is sightless to assign a fundamental role in evolution to repetitive sequences (Fig. 1), as has often happened (Doolittle and Sapienza, 1980; Orgel and Crick, 1980).

Over the next several decades, from the first discovery of TEs (McClintock, 1950), it became apparent that not only do TEs 'jump', but they are also found in almost all organisms (both prokaryotes and eukaryotes) and typically in large numbers (Gierl et al., 1989; Hershberger et al., 1995; SanMiguel et al., 1998; Borodulina and Kramerov, 1999; Tavakoli and Derbyshire, 2001; Yu et al., 2004; Feschotte, 2008; Pace et al., 2008; Stoye, 2012; Lisch, 2013; Oliver et al., 2013; Bourque et al., 2018; Neumann et al., 2019; Gill et al., 2021). For example, TEs make up approximately 50% of the human genome and up to 90% of the maize genome (Wang et al., 1998; International Human Genome Sequencing Consortium, 2001; Zhao et al., 2018; Guio and Gonzales, 2019).

Notably, the possibility that TEs can move within the genome between individuals of the same species and between genomes of phylogenetically very distant species makes any conceptualisation of the individual genome as the only repository of genetic information obsolete (Fig. 1). It is therefore conceivable that the exchange of genetic information could take place between very different levels of genomic organisation (Fig. 1). Thus, although insertions within genes are often deleterious, there is no doubt that TEs can play important roles in evolution in many ways (Volff, 2006; Feschotte, 2008; Pace et al., 2008; Lisch, 2013; Oliver et al., 2013; Ayarpadikannan and Kim, 2014; Chuong et al., 2017; Bourque et al., 2018; Cosby et al., 2019; Koonin et al., 2020; Almojil et al., 2021).

3. Transposition mechanisms

An extensive classification of TEs and their mobilisation systems is widely described (Lisch, 2013; Martin, 2018; Bourque et al., 2018; Neumann et al., 2019; Koonin et al., 2020; Orozco-Arias et al., 2019; Usai et al., 2020b). With an extreme simplification, based on the intermediary used for transposition, TEs can be divided into two main classes: Class 1 (REs) and Class 2 (DNA TEs). Both types of mobile elements are divided into different orders and lineages, basing on sequence homology and on whether or not the TEs encode their own transposition machinery (Fig. 1). The most abundant RE order in plant genomes are long terminal repeats REs (LTR-REs), which are made of the

sequences encoding the transposition machinery, flanked by two long terminal sequences, which are identical at the time of insertion. LTR-REs are generally 5–7 kb long but can even reach lengths of 15 kb, mainly due to the length of the LTRs that can reach several hundred base pairs (Neumann et al., 2019; Fambrini et al., 2020). The internal region contains sequences corresponding to the *group-specific antigen* (*gag*) and *polymerase* (*pol*) genes, which are homologous to those of retroviruses. These genes encode the proteins necessary for retro-transcription: GAG, protease (PROT), integrase (INT), reverse transcriptase (RT), and ribonuclease H (RNAse H) (Defraia and Slotkin, 2014; Bourque et al., 2018; Neumann et al., 2019; Orozco-Arias et al., 2019; Usai et al., 2020b). LTR-REs are divided into two main lineages, *Ty1-Copia* and *Ty3-Gypsy* which, in turn, can be classified into several major evolutionary lineages (Kumar and Bennetzen, 1999; Bennetzen, 2005; Wicker et al., 2007; Llorens et al., 2011; Huang et al., 2012).

Non-LTR-REs lack LTRs. Non-LTR-REs are divided into two classes: long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) (Bourque et al., 2018). LINEs contain a gene encoding an RT (Doucet et al., 2015); conversely, SINEs are short repeated sequences ranging in size from 100 to 600 bp (Schmid and Maraia, 1992; Deragon et al., 1994). SINEs are non-autonomous elements that mobilise using the autonomous enzymatic mechanism of LINEs (Sheldock and Okada, 2000; Dewannieux et al., 2003; Finnegan, 2012; Bourque et al., 2018; Nishiyama and Ohshima, 2018; Fambrini et al., 2020).

REs are mobilised by a reverse transcription process with the formation of an intermediate RNA. The extra DNA copy is then inserted at a different site in the genome. Due to their transposition system, REs have largely contributed to the increase of total genomic DNA in plants and animals, such as *Arabidopsis*, wheat, barley, corn, fig, sunflower, and humans, making up the most represented TEs (Tsay et al., 1993; Feschotte and Pritham, 2007; Finnegan, 2012; Giordani, et al., 2014; Badouin et al., 2017; Mascagni et al., 2015; 2017; Chen et al., 2019; Usai et al., 2020a). Therefore, in most plants and animals, including humans, most of the genome is made up of TEs. In addition, for this transpositional mechanism, RE restructuring is frequent, and RE insertions are responsible for both gene and chromosomal mutations (Fig. 1), as well as changes in gene expression (Bourque et al., 2018; Fambrini et al., 2020). The selection of several orange (*Citrus sinensis*) varieties is one example. In orange, the *Ruby* gene encodes for a MYB transcription factor involved in the regulation of anthocyanin biosynthesis. In the Navalina

variety, the *Ruby* gene is normally functional and shows a rather limited relative expression in the fruit pulp. In the Tarocco variety, the *Ruby* allele shows a RE inserted upstream of the coding sequence (Butelli et al., 2012). The LTR sequence provided a new promoter, resulting in increased *Ruby* expression with consequent accumulation of anthocyanins in the pulp. Recombination between the LTRs also led to a further enhancement of the transcriptional activity of the *Ruby* gene, giving rise to the Maro variety, with increasing concentration of anthocyanins in the pulp (Butelli et al., 2012).

Class 2 elements are classified into two subclasses and different families according to their sequence, terminal inverted repeat (TIR), target site duplication (TSD), and transposition process. Some of the subclass I families enclose *Tc1/mariner*, *PIF/Harbinger*, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *piggyBac*, and *CACTA* (Fambrini et al., 2020). *Helitron* and *Maverick* TEs belong to subclass II, as they perform a distinct process of transposition and insertion into DNA (Morgante et al., 2005; Wicker et al., 2007; Muñoz-López and García-Pérez, 2010; Bourque et al., 2018). Among the non-autonomous Class 2 elements, the miniature inverted-repeat transposable elements (MITEs) are the most common in many eukaryotes, especially in plant species. These short elements (80–500 bp) lack the gene encoding the transposase (Jiang et al., 2004; Bourque et al., 2018). MITEs are flanked by TIR and TSD sequences, and their transposition is mediated by transposases related to autonomous DNA TEs of the same families. In addition, MITEs are frequently associated with alternative splicing events (exonisation) (Jang et al., 2004).

Most Class 2 TEs or DNA TEs can operate insertions and excisions in the host genome in a completely autonomous way by virtue of a specific enzyme, the transposase (Wicker et al., 2007; Finnegan, 2012; Oliver et al., 2013; Stuart et al., 2016; Bourque et al., 2018; Dubin et al., 2018; Neumann et al., 2019; Orozco-Arias et al., 2019). Transferring the DNA TE from its current location to a new destination in the genome employs a ‘cut and paste’ mechanism (Fig. 1). Acting like scissors, the transposase makes three cuts in the host's double helix DNA. Of these, two cuts are borne by the same ends of the TE, which include some terminal repetitive sequences, while a third cut is borne by the target DNA. In this case, the transposition mechanism is autonomous since the DNA regions of the TE encode the transposase. These are the repeated sequences present at the end of the TE, which, acting as recognition and anchoring sites for the same transposases, define the position of the cuts. The ultimate result of this operation is the formation of a synapse between the two ends of the TE and the interaction between the respective transposases:

a necessary and sufficient condition for the formation of an active complex capable of cutting the double strand of the DNA TE. Once excised from the original site, the TE is then transferred to another region of the host DNA or even to another chromosome by the ability of the transposases to localise, recognise, and bind to a new target site. Even TEs that have lost all or part of their transposase (non-autonomous) can be mobilised (Fedoroff et al., 1983). To perform the transposition, non-autonomous elements rely on enzymes encoded in *trans* by autonomous elements (Sabot and Schulman, 2006; Fambrini et al., 2011; 2014a).

This mechanism is called conservative because it allows the transfer of a nucleotide sequence from one region of the DNA to another without any copy remaining in the original site. Therefore, in the absence of replication, the TE is effectively eliminated from the chromosome or plasmid of origin. However, excision frequently generates genetic 'fingerprints' that can cause variability in the old insertion site (Fambrini et al., 2011). Frequently, multiple allelism phenomena are generated by these imprecise excision mechanisms (Fambrini et al., 2014a).

In many prokaryotes, there are DNA TEs that are mobilised in the genome by a transposition that is called replicative. In this case, during transposition, a new copy of the transposable sequence is generated, which is why, in addition to the one transferred to the new site, another copy that remains in the original site, with easily imaginable consequences on genomic amplification, is also formed (Tavakoli and Derbyshire, 2001). TEs using this mechanism are known as composite TEs that consist of two inverted repeats from two separate TEs moving together as one unit and carrying the DNA between them. For example, consider a segment of DNA flanked at both ends by two identical insertion sequences. The transposase will move any segment of DNA surrounded by a pair of inverted repeats that it recognises. Consequently, when transposition occurs, there are several possibilities. First, each of the insertion sequences may move independently. Second, the whole structure between the two outermost inverted repeats may move as a unit, that is, as a composite TE. Replicative transposons need an extra enzyme, resolvase, and a site in the DNA for it to act on. Many of the well-known bacterial TEs that carry genes for antibiotic resistance or other properties are composite TEs (Sun et al., 2019).

4. Evolutionary impact of transposition

4.1. The complex living system

To understand the mechanisms of action of the TEs, it is necessary to refer to the concept of information according to which everything that a living organism achieves on a physiological level is attributable to the action of its genes (Dawkins, 1976; Doolittle and Sapienza. 1980). At the same time it will be important to contextualize the activity of the TEs in the evolution of a complex living system.

In the reductionist view, it is implicitly assumed that the structures and functions that the organism expresses are causally correlated with the sole functionality of the genetic program encoded in its DNA. The reductionist approach aims to decrypt phenotypic variability in stages, founded on the underlying hypothesis that genome-to-phenotype relations are largely constructed from the additive effects of their molecular players (Williams and Auwerx, 2015). Every event is attributable exclusively according to the strictly logic of the principle of causality (Monod, 1970). The most significant consequence of this assumption is that natural selection operates on the genes that the organism carries (Dawkins, 1976), rather than on the organism as the subject of natural selection (Darwin, 1859). The reductionist method dissecting biological systems into their constituent parts has been effective in explaining the biochemical and molecular basis of numerous living processes in animals and plants. However, this approach has some limitations. Gilbert and Sarkar (2000) argued against reductionism, as it insufficiently explains phenotypes and is unable to predict development. Indeed, the reductionist program has not uncovered the fundamental laws from which we can deduce how a particular living system actually works (Cohen and Harel, 2007). As an example of why the reductionist approach fails, consider the function of one cell within a multicellular organism. Even if we understand the cell's function, that does not mean we fully understand the organism's physiology. In 1970, already Jacob proposed a non-reductionist approach to study biological phenomena, introducing the concept of ‘integron’ (Jacob, 1976). This concept, not to be confused with some type of TEs (Domingues et al., 2012), is similar to what is understood today when speaking of complex living systems. In fact, Jacob (1976) claims that the advances of molecular biology mark, at last, the merging of physiology and natural history: organisation and evolution, molecular mechanisms of heredity and teleonomy can finally be treated in a unified biological framework. As a consequence, the nested organisation of living beings, composed of layers of integrated sub-units (called integrons), becomes a consequence of

evolutionary processes. Biological systems are extremely complex and have emergent properties that cannot be explained or even predicted by studying their individual parts (Van Regenmortel, 2004). The systems approach aims to examine large-scale interactions of many components simultaneously, on the premise that interactions in gene networks can be both linear and nonlinear (Williams and Auwerx, 2015). The reductionist approach underestimates this complexity and therefore can have an increasingly detrimental influence on many areas from physical, biological, and physiological research. As Gould and Eldredge (1993) wrote, the "*most important implications of punctuated equilibrium remain the recognition of stasis as a meaningful and predominant pattern within the history of species, and the recasting of macroevolution as the differential success of certain species (and their descendants) within clades.*" Thus, these authors rejected the reductionist view that natural selection operates exclusively on discrete genes but also the idea that selection on individuals in populations is the sole force shaping large-scale evolutionary trends.

Complex biological activity does not arise from the specificity of the individual components that are involved (Strogatz, 2000; Van Regenmortel, 2004; Cohen and Harel, 2007; Aziz and Caetano-Anollés, 2021). Thus, a systems approach favours the concept of the reciprocal relationships between genes, organisms, and their environment, in which all three elements act as both causes and effects (Alm and Arkin, 2003). In addition, it emphasises the selective advantage of these features during evolutionary history. Functional explanations are more useful for understanding complex biological systems with many interactions than are causal explanations that give unwarranted importance to a single factor (Van Regenmortel, 2002).

According to a systems approach that is conceptually opposed to the reductionist one, the relationship that exists between form and information in a living organism must be conceived as an emerging dimension with respect to the structural conditions that precede it. Indeed, a living organism is an active system and not mere sequential transformations of information encoded by a DNA coding. Living cells, organisms and species emerge from the concurrent interactions of various entities and processes; DNA is only one of the many participating informational entities (Cohen and Harel, 2007). The activation of specific genes, includes TEs, emerges from the dynamic state of the cell, which interacts, among others, with environment, protein expression, developmental cell phase, epigenetic marks (Fig. 2). Therefore, the comparison between the reductionist and systems approaches

359 requires us to clarify what is meant by information and what role TEs could play in its
360 implementation during evolution of cells and living organisms.

361 In fact, every living system has a digital coding DNA (or RNA) capable of providing a
362 kind of self-referential description (genotype) through the combinatorial use of individual
363 letters (Pattee, 1972), which eventually contributes to germ cells. On this view, any DNA
364 gene is a carrier of digital information by virtue of its unique base sequence (Janga et al.,
365 2009; Travers et al., 2012; Muskhelishvili and Travers, 2013). However, in a living system
366 the DNA also provides an analog information to express this same description in both
367 structural and functional terms (phenotype). While the digital information uses only
368 combinatorial rules of a order to build a non-temporal memory system, the analog
369 information makes these same rules significant every time they update them in the
370 temporal domain (Wilden, 1980). Although there is no definitive answer, it is still possible
371 to speculate on the evolutionary advantages that the use of discrete codes could offer
372 compared to analog dynamics alone. In the absence of discrete coding, the system would
373 necessarily have been autopoietic, and each component would have had to catalyse its
374 own formation through a catalytic synthesis process (Hordijk and Steel, 2015). If that had
375 been the case, evolution would have proceeded much more slowly, and the survival of the
376 fittest would have been threatened by a higher likelihood of copying errors. The benefits of
377 discrete genetic coding must therefore be sought in the possibility of constructing arbitrary
378 and improbable combinations and of making identical heritable copies of them in forms
379 completely separate from the need to associate them with catalytic functions (Cariani,
380 1998).

381 DNA information is expressed in the analogical form when the nucleotide sequence
382 is translated for the synthesis of proteins, while it is replicated in digital form when it is only
383 handed down unchanged from one generation to another through a semi-conservative
384 replication mechanism. To fully understand what is meant by information in this context, it
385 is necessary to show the causal connection that exists between the nucleotide sequence
386 of DNA and the series of phenotypic structures that its reading produces. The sequence
387 itself cannot be the cause of any biological event. Natural selection operates onto
388 phenotypes; therefore, the so-called ‘survival of the fittest’ still refers to the organism as a
389 unit and not to any of its parts. The outcome of this struggle for survival is natural selection
390 (Cohen, 2016). For Darwin, any individual plants and animals that happened to vary in an
391 advantageous way would be more likely to triumph over their competitors. Only the

survivors would produce offspring, which might diversify and develop further to fill any available ecological niche (Darwin, 1859). This follows that the knowledge of what survives does not give us any opportunity to understand the mechanisms through which genotypic information is mapped on the respective phenotypic characteristics during evolution. Mapping indicates the correspondence between the domain of the nucleotide sequences of the DNA and the real characteristics of the organisms.

In molecular biology, the use of the concept of information refers exclusively to the computational theory of communication according to which information (entropy) is measured by the logarithm of the number of available choices (Shannon and Weaver, 1964). In this perspective, the information coincides with the fidelity with which a message can be communicated and is therefore a quantitative measure of the improbability with which it reaches a recipient. According to this interpretation, the information depends only on the statistical distribution of the symbols (nucleobases) of which a nucleotide sequence is composed (Gentlemen and Mullin, 1989).

The most important result that this definition has allowed us to achieve is the comparison between different genomes through the alignment of homologous sequences. However, this procedure still tells us nothing about the meaning that emerges from the comparison between the variety of sequences found in the different organisms and the relative phenotypic characteristics for which they encode. In other words, adherence to this restrictive concept of information entails the loss of its semantic content and consequently the lack of an understanding of the relationship between nucleotide sequences and the system that interprets them as instructions for its own construction (Wills, 2014).

In essence, it is no longer a question of considering the cellular structure as an ultimate product of the translation process of genomic information. In contrast, the cellular structure and the genome must be seen as partners in a constant homeostatic interaction, one corresponding to the temporal realisation of the analogical code and the other including its digital memory (Hoffmeyer and Emmeche, 1991; Hoffmeyer, 2008). The fact that natural selection favoured the persistence of the relationship between replication and gene translation, on the one hand, and related protein products, on the other, suggests that both processes might have been compartmentalised within shared spaces from the beginning of life (Wills, 2016). In addition, it is very likely that it was the same co-localisation that favoured the emergence of reaction-diffusion mechanisms between polymeric phenotypes and the production of autocatalytic cycles (Kauffman, 1986).

Moreover, when a set of molecules is confined within a shared space, the spontaneously created autocatalytic cycle leads to the formation of a self-assembly system within which each component is automatically connected to the others over time (Rodon Fores et al., 2020). The emergence of a spontaneous relationship is such that only those polymeric sequences that express it can be selected for the construction of a code of correspondence between structure and function (Wills, 2001; Hordijk and Steel, 2018). This is an essentially computational process that, in the long run, allows for the creation of new meanings between the memory system and the polymeric elements that have acquired a role for the replication and translation functions that they realise (Carter and Wills, 2017). TEs probably played a role in realizing the correlation between digital and analog information with transcriptional and post-transcriptional epigenetic mechanisms. This is to allow the living system to explore new relationships for each newly developed combination (Fig. 2).

4.2. Co-evolution between transposable elements and their hosts

Given the widespread presence of TEs in the genome of almost all living organisms, including mammals, it is natural to presume that their persistence in the host represents an evolutionary advantage and that, for this reason, it could have some benefit for the host (Volff, 2006; Feschotte, 2008; Pace et al., 2008; Lisch, 2013; Oliver et al., 2013; Ayarpadikannan and Kim, 2014; Chuong et al., 2017; Bourque et al., 2018; Cosby et al., 2019; Koonin et al., 2020; Nishihara, 2020; Almojil et al., 2021; Gill et al., 2021). It is based on the considerations that some authors have proposed to consider the genome as equivalent to a “habitat” or “ecosystem” that TEs colonised over time and from which they were then domesticated (Kremer et al., 2020). This trend is clearly evidenced by numerous comparative studies (Habibi et al., 2015; Bourque et al., 2018), according to which the presence of TEs enhances with increasing biological complexity while, at the same time, the percentage of the genome occupied by coding genes decreases (Boeke and Devine, 1998).

However, to verify the effects that TEs have exerted in the long term, it is necessary to evaluate their persistence in relation to those mechanisms, which, on the one hand, guarantee their replication and, on the other, suppress their activity. If the genome can be compared to a “habitat” (Kremer et al., 2020), the TEs that have settled there can be

457 considered parasites. As such, they had to deal with two different strategies: either (1)
458 making the most of the host's resources with the risk of cause their own elimination or (2)
459 co-evolving to build a lasting symbiotic relationship (Schrader and Schmitz, 2019).

460 In the first case, the TEs must guarantee a high level of proliferation within the
461 genome, keeping their proliferation rate below a threshold beyond which they would be
462 detrimental to their host. Faced with excessive proliferation and a consequent risk for the
463 survival of the host, TEs can still make use of the alternative strategy of starting a new
464 cycle of proliferation through horizontal transfer, virtually affecting all major families of TEs
465 in all phylogenetic branches of the tree of life (Wright and Voytas, 1998; Diao et al., 2006;
466 Gao et al., 2014; Hosaka and Kakutami, 2018; Ivancevic et al., 2018; Thybert et al., 2018;
467 Wallau et al., 2018; Nishihara, 2020). It is the availability of a transfer vector (virus,
468 bacteria, nematode, or other parasites) that allows TEs to mobilise between genomes of
469 phylogenetically distant species (Walsh et al., 2013; Gao et al., 2014; Bourque et al., 2018;
470 Gilbert and Feschotte, 2018) (Fig. 1).

471 Some believe that horizontal transfer has a major influence on the speciation process
472 of placental mammals (Chuong, 2013). The separation of eutherians from monotremes
473 (Warren et al., 2008), which occurred about 160–190 million years ago, most likely
474 coincides with the invasion of LINEs into the genomes of placentates (Waters et al., 2007).
475 Genomic expansion and the increase in genetic variability caused by LINE-1 elements
476 may have allowed the mammalian genome to adapt to multiple environmental conditions
477 and regulate gene transcription in relation to the completion of foetal development via the
478 placenta (Ivancevic et al., 2016; 2018; Rodriguez-Terrones and Torres Padilla, 2018). In
479 contrast, the lack of genetic variability limited the evolutionary expansion of monotremes,
480 forcing them to adapt to smaller ecological niches.

481 To co-evolve with the host, TEs must have derived mechanisms to ensure their
482 permanence in the genome as tolerated parasites. For example, in humans, there are
483 about 500,000 copies of the LINE-1 element (Sassaman et al., 1997). However, regardless
484 of the mode of transmission, whether vertical within the same species or horizontal
485 between different species, there is no possibility for TEs to have a guaranteed perpetual
486 survival (Blumenstiel, 2019). Presuming persistence in the host as the result of a
487 precarious balance between the level of proliferation and the incidence of mutations, it is
488 easy to understand how a low transposition rate leads to the extinction of the transposable
489 sequences. The most significant result of this condition is the continuous erosion of TEs by

point mutations, which are prevented from producing new integration events due to the loss of coding for the transposition enzymes. For example, more than 99.9% of the 500,000 copies of LINE1 elements in the human genome have undergone various forms of structural mutations and arrangements, so that only about 100 are still active (Nishiyama and Ohshima, 2018). In summary, TEs owe their evolutionary success, both horizontal and vertical, to a condition of equilibrium between transposition and selection, as it has gradually been defined over the course of evolution (Fig. 2).

The relationship between TEs and their hosts is therefore attributable to a competitive/cooperative relationship and has induced the transposable sequences to change over time due to the accumulation of point mutations or genomic integration (Bourque et al., 2018; Nishihara, 2020). Stochastic shuffling by TEs is most likely due to the imprecision with which they are inserted and/or removed from the genomic content of the host (Venner et al., 2009). Due to this inaccuracy, transposition processes can indirectly affect the adjacent sequences of the host genome, including coding genes and regulatory sequences (Jiang et al., 2004). However, the existence of a competitive relationship with the genome makes the transposition more complex than it might appear upon first analysis.

Regardless of the type of element, in evolutionary terms, the insertion sites of TEs are truly significant (Fig. 1). Furthermore, the genetic consequences can be very different if the TEs are inserted in the coding region of a gene, in proximity to a gene (e.g., promoter region), in introns or untranslated regions of the gene, and in heterochromatic regions of the chromosome (Greene et al., 1994; Wicker et al., 2007; Finnegan, 2012; Oliver et al., 2013; Bennetzen and Wang, 2014; Stuart et al., 2016; Bourque et al., 2018; Dubin et al., 2018; Neumann et al., 2019; Orozco-Arias et al., 2019; Fambrini et al., 2020). In contrast to the expectations, the distribution of TEs is not random but shows a clear trend towards confinement in preferential compartments of the genome (Capel et al., 1993; Wicker et al., 2007; Finnegan, 2012; Bourque et al., 2018). As a result of opposing selective forces, TEs tend to occupy those regions of the genome that guarantee their maximum proliferation, while the host will tend to confine them to regions that mitigate their deleterious effects (Almojil et al., 2021). Indeed TEs can sometimes be advantageous at the species level as a source of evolutionary innovation capable of modifying existing genes or originating novel genes through the mechanism of exaptation (Joly-Lopez and Bureau, 2018), a process whereby TEs conferring positive phenotypic benefits to the host genome begin to

evolve as conventional genes rather than through self-duplication (Volff et al., 2006; Tsuchiya et al., 2013; Hoen and Bureau, 2015). TEs exaptation provided a major contribute to the evolution of gene regulation (De Souza et al., 2013) and indeed in plant genomes were found 'exapted transposable element genes' (ETEs) involved in physiological functions such as flowering (Cowan et al., 2005), development (Bundock and Hooykaas, 2005), light signaling (Hudson et al., 2003) and stress response (McClintock, 1984; Chénais et al., 2012; Makarevitch et al., 2015).

TEs can also alter gene expression by inserting within promoter or intronic regions (Umeda et al., 1991; Selinger and Chandler, 2001; Roccaro et al., 2005; Fambrini et al., 2014a), inactivate genes by inserting into exonic regions, and causing extensive genomic restructuring (Fambrini et al., 2011; 2018; Zhang et al., 2011; Chapman et al., 2012) (Fig, 1). Conversely, the host will tend to minimise the deleterious impact of the transposition and to react by favouring insertion in less transcriptionally active parts of the genome, such as heterochromatic and centromeric regions (Gao et al., 2008). Therefore, the mobilisation of TEs during evolution and early embryogenesis may have contributed to the construction of mosaic genomes, including highly sophisticated regulatory networks (Garcia-Perez et al., 2016). In addition, TE-derived proteins have been recurrently repurposed as part of defence systems that protect prokaryotes and eukaryotes against the proliferation of infectious or invasive agents, including viruses and TEs themselves (Jangam et al., 2017) (Fig.1).

It is possible to distinguish the mutations that confer new functions to the genome (gain-of-function) from those that instead involve a loss-of-function. This distinction could also be applied to mutations induced by TEs (Wicker et al., 2007; Finnegan, 2012; Oliver et al., 2013; Stuart et al., 2016; Bourque et al., 2018; Dubin et al., 2018; Neumann et al., 2019; Orozco-Arias et al., 2019). Frequently, the insertion of a TE in the coding region of a gene induces a null mutation (Fambrini et al., 2011; Chapman et al., 2012; Ventimiglia et al., 2020). However, the insertion of a TE outside the coding region can also interfere with the activity of the promoter, modifying the regulatory regions or the transcription start sites, phenomena that alter or abolish transcription (Dubin et al., 2018; Neumann et al., 2019; Orozco-Arias et al., 2019; Fambrini et al., 2020). Nevertheless, there are also examples of TE insertions, especially in promoter regions, which cause a gain-of-function for the involved genes (Roccaro et al., 2005; 2007; Chapman et al., 2012; Fambrini et al., 2014a; 2014b; 2018).

This could explain the scenario proposed for the origin of eukaryotic chromosomes. In eukaryotic chromosomes, the physical ends of chromosomes are telomeres. The maintenance of telomeres requires telomerase, a specific RNA-dependent DNA polymerase enzyme complex that carries its own RNA template and adds telomeric repeats to the ends of chromosomes using a reverse transcription mechanism. In *Drosophila*, LINEs form and maintain telomeres to replace the telomerase enzyme lost during Diptera evolution (Gladyshev and Arkhipova, 2007; Pardue and DeBaryshe, 2011). Phylogenetic analysis of *Drosophila* telomeres demonstrated that multiple autonomous and non-autonomous ancient REs were recruited to perform telomere maintenance (Villasante et al., 2007; Pardue and DeBaryshe, 2011; Casacuberta, 2017), a mechanism that is considered a primary evolutionary mode of protection for linear chromosome ends. This molecular domestication event of ancient TEs could be seen as a replica of what could have happened much earlier in the evolution of eukaryotic chromosomes. This solves the problem created by the linearisation of chromosomes from a circular ancestral chromosome after a double strand break—sticky endings with deleterious consequences in the mitotic process (Belfort et al., 2011; Fulcher et al., 2014; Kordyukova et al., 2018). The transition of an ancestral circular genome to multiple linear chromosomes was crucial for the development of eukaryotic organisms because it allowed rapid evolution from aneuploid to polyploid ancestral genomes (Garavís et al., 2013). In a theory elaborated by E. V. Koonin (2006), the origin of linear chromosomes is associated with the invasion of archaeal hosts by an α-proteobacterial progenitor that resulted in mitochondrial endosymbiosis and the invasion of Group II self-splicing introns (Lambowitz and Zimmerly, 2004), the most ancient genetic entities (Garavís et al., 2013). Group II introns were suggested as eukaryotic evolutionary ancestors of REs and spliceosomal introns. They consist of a catalytically active intron RNA and an intron-encoded RT that is related to non-LTR REs and assists splicing by stabilising the catalytically active RNA structure (Eickbush and Jamburuthugoda, 2008; Lambowitz and Zimmerly, 2011). It is believed that the attachments of non-LTR REs at the ends of DNA double-strand breaks have finally developed to ‘proto-telomeres’ of the initial linear chromosomes (Garavís et al., 2013). Therefore, telomere function and TE regulation seem to be evolutionarily joined (Servant and Deininger, 2015; Kordyukova et al., 2018). Moreover, telomere elongation, as well as TE activation, occurs in the narrow time window during gametogenesis and early development, suggesting that common principles could drive both processes (Richardson et al., 2017).

5. Epigenetic regulation of transposable elements

The development of each cell results from the combination of a two logically distinct types of information provided by the DNA: digital, which represents its specific memory to be handed down unchanged from generation to generation, and analog that expresses its functionality in the temporal domain (Hoffmeyer, 2008). Therefore, the nature of the contribution of TEs as major component of large organism genomes and main drivers of genome evolution is evident (Fig. 1). In fact, the insertion of new sequences and or mobilisation of TEs has made it possible to expand the genome complexity and extend the range of possible combinations (Gregory et al., 2007; Lee and Kim, 2014; Bourque et al., 2018; Fambrini et al., 2020). For example, different stressors, such as high and low temperatures, and interspecific hybridisations can influence TE mobilisation (Grandbastien, 1998; Lang-Mladek et al., 2010; Pecinka et al., 2010; Tittel-Elmer et al., 2010; Ito et al., 2011; Bartlett and Hunter, 2018; Dubin et al., 2018; Rech et al., 2019). This is not surprising, considering that some TE families contain sequences that constitute stress response elements (SREs) in their own promoters and are sensitive to external signals, such as biotic and abiotic stress (Cavrak et al., 2014; Wibowo et al., 2016; Dubin et al., 2018; Vangelisti et al., 2019).

Given the extent of climate change recorded over the geological eras, it is very likely that this strategy has been instrumental in favouring the expansion of the adaptive solutions of many organisms. Additionally, given the irregularity with which periods of rapid change have alternated with long periods of stability during evolution, it is equally likely that TEs have been inserted as predicted by the theory of punctuated equilibria (Eldredge and Gould, 1972). The coincidence between the greater diffusion of TEs and the periods of more rapid change can be thought of as the best solution selectively available to organisms to respond to the stress generated by the change. Neither climate change nor genomic adaptation alone can have causal value, but the relationship between the two does, implying that conditions of balance must exist between the flexibility of the host genome and the proliferative capacity of TEs (Casacuberta and Gonzales, 2013; Gill et al., 2021; Muyle et al., 2021).

Absolute stability would effectively prevent the evolution of hosts and TEs, while excessive flexibility would inevitably lead to their extinction. From this perspective, it is therefore interesting to consider which and how many mechanisms organisms have been

able to devise to respond to the challenge represented by the diffusion of TEs, and how these same mechanisms have allowed them to evolve towards levels of increasing structural and functional complexity (Ali et al., 2021). In other words, if TEs have provided the prime elements for the construction of a combinatorial syntax, the adaptive response of organisms has provided semantic content to this growing digital grammar. One hypothesis suggests that TEs have caused the genome to evolve towards levels of greater complexity and to create new genes and increasingly complex gene regulation networks (Sundaram et al., 2014; Rayan et al., 2016; Vicient and Casacuberta, 2017; Trizzino et al., 2018).

Despite these concepts, an arms race has been established between TEs and hosts, and several mechanisms appear to have emerged during eukaryotic evolution to suppress TE activity (Fig. 1). However, the consequence of the struggle between TEs and hosts should be considered as a passage from a first phase characterized by the invasion of TEs welcomed by the host as parasites, up to the evolution of mutualistic relationships between TEs and hosts (Cosby et al. 2019).

An assortment of host-encoded mechanisms are known to repress eukaryotic TEs. Most often, it involved epigenetic mechanisms relating to a variety of pre-transcriptional to post-translational events (Fedoroff et al., 1995; Slotkin and Martienssen, 2007; Song and Schaack, 2018; Deniz et al., 2019; Drongitis et al., 2019). Among the mechanisms most frequently used by host genomes to counter their transposition are the methylation of DNA cytosines, the modification by deacetylation of histone proteins, and silencing by microRNA (miRNA) directed towards complementary transposable sequences and small interfering RNA (siRNA) for RNA-directed DNA methylation (Law and Jacobsen, 2010; Ito et al., 2011; 2013; Li et al., 2015; Zhang et al., 2018; Jansz, 2019; Koonin et al., 2020). Overall, these epigenetic mechanisms tend to make most of the TEs dormant or inactive, with the dual objective of repressing their activity and inducing a progressive compaction of the chromatin of the adjacent regions (Oliver and Greene, 2009). Notably, this implies that the compacted state of the adjacent chromatin can extend to the promoter of the proximal gene and therefore suppress its expression (Dubin et al., 2018).

The presence of methyl groups on DNA molecules was discovered by Wyatt in 1951. The methylation of DNA directly silences transcription. In animals, DNA methylation regards cytosines placed in the dinucleotide CpG, but in plants, the CpHpG and CpHpHpG patterns (where H = A, T, or C and p is a phosphate group) can also be of interest. Among these modifications, 5-methylcytosine (5mC) and N4-methylcytosine (4mC) are the most

studied. To date, these modifications can be predicted through three main approaches; chemical conversion-based methods, enzymatic-based methods and direct DNA-based methods (Lv et al., 2021). In mammals, DNA methylation, catalysed by DNA methyltransferases (DNMTs), can be reversed by the ten-eleven translocation (TET) proteins (Wu and Zhang, 2010). In plants, the regulation of DNA methylation is, to some extent, independent of the different sequence contexts (Law and Jacobsen, 2010). For GC sites, DNA methyltransferase (MET1) catalyses the transfer of a methyl group to cytosine to maintain the epigenetic state during DNA replication. Cytosine-5 DNA methyltransferases (C5-MTases) mainly accomplish two functions: recognition of a specific DNA sequence and catalysis of the transfer of a methyl group from the cofactor S-adenosyl-L-methionine (AdoMet) to carbon 5 in the pyrimidine ring of cytosine residues (Kumar et al., 1992). CHG sites are methylated by CHROMOMETHYLASE3 (CMT3), depending on the chromatin state of dimethylation of lysine 9 of histone H3 (H3K9me2) (Law and Jacobsen, 2010; Zemach et al., 2010; Nicetto and Zaret, 2019). Since CHH methylation does not have a template for maintenance, it must be methylated *de novo* after replication. Asymmetric CHH methylation is maintained by *de novo* methylation, which is catalysed by CMT3 and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), an ortholog of the mammalian Dnmt3a/b *de novo* methyltransferase (Chan et al., 2005). DRM3, a catalytically mutated DNA methyltransferase paralog, also plays an important role in establishing *de novo* cytosine methylation in all sequence contexts in the process of RNA-directed DNA methylation (RdDM) by stimulating the activity of DRM2 (Henderson et al., 2010; Saze et al., 2012). The average level of cytosine methylation in vertebrates is 3–8% of all cytosine, but the highest levels among eukaryotes have been observed in plants, usually 6–30% (Chen et al., 2004), and up to 50% in some plants, e.g., maize (Suzuki and Bird, 2008). Such high methylation in plants is due to the methylation of both CpG and CpHpG sequences while in animals, methylation usually only concerns cytosines located in CpG dinucleotides (Pradhan and Adams, 1995). The presence of DNA methylation in promoter regions may prevent RNA polymerase from binding *via* steric hindrance, thereby inhibiting transcription. The removal of the methyl group from cytosine (demethylation) can proceed in a passive or active way. In *Arabidopsis*, the DEMETER (DME) family of DNA glycosylase actively removes DNA methylation by a base-excision repair (BER) mechanism (Choi et al., 2002; Gong et al., 2002; Agius et al., 2006) and thus has a key role in the plasticity of the epigenome (Zhu et al., 2007; Zhu, 2009).

689 In a wide range of plants, genes and TEs adopt distinct patterns of methylation. Plant
690 genes tend to have low levels of methylation at CHG and CHH sites within the coding
691 regions, while different patterns have been observed in CG sites, which can exhibit
692 different ratios between exons and intron methylation levels (Feng et al., 2010; Bucher et
693 al., 2012; Finnegan, 2012; Bourque et al., 2018; de Mendoza et al., 2018; Usai et al.,
694 2020a). CHG methylation is rare in exon regions; conversely, the occurrence of extensive
695 methylation is common in introns (Zhang et al., 2006; Zilberman et al., 2007; Cokus et al.,
696 2008; Rigal and Mathieu, 2011; Bucher et al., 2012; Rigal et al., 2016; Zakrzewski et al.,
697 2017); in many instances, these introns contain a complete or defective TE (Roudier et al.,
698 2011; Du et al., 2012; Hirsh and Springer, 2017; Zhang et al., 2018). Maintaining
699 methylation in non-CG sites of heterochromatic TEs also requires the activity of CMT2
700 (Zemach et al., 2013). Lysine 9 methylation of histone H3 (H3K9me2) is mediated by the
701 SET domain of SU (VAR) 3-9 HOMOLOGUE 4/KRYPTONITE (SUVH4/KYP), SUVH5, and
702 SUVH6 proteins, which promotes the binding of CMT3 to chromatin and maintains
703 methylation in CHG sites (Law and Jacobsen, 2010; Du et al., 2012). Furthermore, the
704 activity of DECREASE IN DNA METHYLATION 1 (DDM1) is required both to modify the
705 chromatin conformation and to maintain the DNA methylation patterns of the TEs inserted
706 in the heterochromatic regions, which are distinct targets of the RNA-directed DNA
707 methylation (RdDM) pathway (Stroud et al., 2013; 2014; Zemach et al., 2013).

708 The *ddm1* mutant of *Arabidopsis* is unable to methylate DNA in centromeric repeats
709 (Vongs et al., 1993). These mutants have become a useful tool for studying the role of
710 epigenetic silencing and activation processes in specific plant TE classes (Gendrel et al.,
711 2002; Lippman et al., 2003; 2004). An explicative example is the *FLOWERING*
712 *WAGENINGEN* (*FWA*) locus of *Arabidopsis*: in this circumstance, a SINE element,
713 immediately upstream of the *FWA* coding region, was epigenetically silenced in the
714 vegetative tissues (Soppe et al., 2000). This process involved small RNAs and DNA
715 methylation (Kinoshita et al., 2007). The ectopic expression of *FWA* in *ddm1* causes a late
716 flowering phenotype. A second epiallele, *bonsai* (*bns*), originated from the *ddm1* mutant
717 after activation of a non-LTR-RE downstream of the *BNS* coding sequence (Saze and
718 Kakutani, 2007). Through genetic selection of mutants, which showed an increase in
719 methylation, in the INCREASE IN BONSAI METHYLATION 1 (IBM1) protein, the
720 JUMONJI domain was also identified (Saze and Kakutani, 2007; Saze et al., 2008). IBM1
721 is a member of the JHDM2/KDM3 family, which includes plant-to-mammal conserved
722 H3K9me demethylases (Klose et al., 2006; Sun and Zhou, 2008; Shen et al., 2014). The

ibm1 mutation induced a combination of developmental abnormalities, which were suppressed by mutants in both *KYP/SUVH4* and *CMT3* genes. The results suggested that the ectopic location of chromatin methylation H3K9me and CHG sequences was responsible for the developmental abnormalities induced by *ibm1* (Saze et al., 2008). The *ibm1* mutation induced extensive hypermethylation in thousands of genes in all chromosomal euchromatic regions. Hypermethylation was specific for CHG sites within genes. Therefore, it has been proposed that IBM1 can protect transcribed genes from CHG methylation and maintain genome integrity by distinguishing between genes and TEs (Miura et al., 2009).

Methylated and silenced TEs are generally excluded from gene regions, suggesting a trade-off between gene expression and TE silencing (Hollister and Gaut, 2009). Despite this, genomic-level studies in eukaryotes showed that there were substantial numbers of intragenic TEs in both animals and plants (Nekrutenko and Li, 2001; Jiang and Ramachandran, 2013; Choi and Purugganan, 2018). In addition, host factors such as IBM2/ANTI-SILENCING 1 (ASI1)/SHOOT GROWTH 1 (SG1) and ENHANCED DOWNY MILDEW 2 (EDM2), which are specifically required for gene transcription in heterochromatic regions, have been identified (Saze et al., 2013; West et al., 2014).

In *Arabidopsis*, the epigenetic modification of a TE can attenuate, rather than eliminate, gene expression. Actually, epigenetic silencing of a TE inserted into the first intron of the *FLOWERING LOCUS C* (*FLC*) gene is associated with a reduced expression of this gene in natural populations, a property that may have been selected for the evolution of late flowering accessions (Liu et al., 2004). It might be expected that the genomes of plants with a large amount of TEs would show more of this type of variation. Furthermore, the rice genome, which is considerably larger than the *Arabidopsis* genome, shows evidence for several TEs that appear to regulate neighbouring genes by antisense transcription. In plants, LTR-REs are generally methylated and transcriptionally inactive (Gehring et al., 2009; Usai et al., 2021). However, non-methylated LTR-REs can become transcriptionally active, resulting in transcriptional activation that extends into the flanking sequences (Kashkush et al., 2003). *Dasheng* is an RE that is polymorphic in numerous rice species and subspecies (Jiang et al., 2002); the expression of some *Dasheng* presents tissue-specific DNA methylation correlating with nearby gene expression tissue specificity (Kashkush and Khasdan, 2007). Furthermore, non methylated copies of

Dasheng REs impact host gene expression by producing antisense chimeric transcripts that putatively promote mRNA degradation (Kashkush and Khasdan, 2007).

Notably, flowering plants have epigenetic control of the spatial distribution of TEs during the fertilisation process. In *Arabidopsis*, the pollen vegetative nucleus undergoes a programmed loss of heterochromatin, resulting in TE activation, TE transposition, and TE siRNA production (Slotkin et al., 2009). These epigenetically activated siRNAs accumulate at higher levels in the neighbouring sperm cells in which TEs are transcriptionally silenced. The authors suggested that the vegetative nucleus sacrifices the integrity of its genome to provide protective siRNAs to sperm cells.

In addition to transcriptional silencing phenomena, post-transcriptional inactivation mechanisms of TEs are also active. RNA interference (RNAi) is one of the main mechanisms by which double strand RNAs (dsRNAs) are cleaved by members of the DICER protein family into short sequences of 21–30 nucleotides that originate in the siRNA that drives the RNAi (Lee et al., 2004; Leuschner et al., 2006). Argonaute (AGO) proteins bind to siRNAs and form the RNA-induced silencing complex (RISC), in which siRNAs recognize the target genes with complementary sequences and AGO proteins cleave and repress the targets (Ghildiyal and Zamore, 2009).

In *Drosophila melanogaster*, a series of studies demonstrated one of the most interesting post-transcriptional control mechanisms for the diffusion of TEs. *P* TEs are one of the best-studied eukaryotic transposons in metazoans (Majumdar and Rio, 2015). These elements were initially discovered in the late 1960's because they cause a syndrome termed hybrid dysgenesis in *D. melanogaster* (Hiraizumi, 1971). Hybrid dysgenesis is the appearance of a set of unusual characteristics in the progeny of crosses between certain strains of *D. melanogaster* (Bucheton et al., 1984). The molecular cloning and biochemical characterization of the P element transposition reaction have led to general insights regarding eukaryotic cut-and-paste-transposition (Bingham et al., 1982; Bucheton et al., 1984). Two independent systems of hybrid dysgenesis, *I-R* and *P-M,* are known, and strains of *D. melanogaster* may be classified as one of two types with respect to each system (Bingham et al., 1982; Bucheton et al., 1984). In the last years, *P*-element-induced wimpy testis (PIWI)-interacting RNAs (piRNAs) were demonstrated to be master repressors of TE activities in *D. melanogaster* (Luo and Lu, 2017; Luo et al., 2020). This small non-coding RNAs identify their targets, mRNAs of active TEs, through perfect or nearly perfect antisense matching (Brennecke et al., 2007). piRNAs repress their target

mRNAs through a “Ping-Pong” cycle (Czech and Hannon, 2016). The interaction between piRNAs and TEs also explained the hybrid dysgenesis for the *I-R* and *P-M* systems (Luo and Lu, 2017). In addition, the population genetic analysis showed the importance of piRNA protection to the host organism. In particular, Luo et al. (2020) suggested that the *de novo* generation of piRNAs is a key mechanism to repress the newly invaded TEs. Thus, it was demonstrated the existence of an evolutionary arms race between the copy numbers of TEs and the abundance of antisense piRNAs at the *D. melanogaster* population level (Luo et al., 2020).

While epigenetic modifications and post-transcriptional/translational suppression may act against the activation of TEs, much less known are the molecular processes involved in the targeted removal of individual TEs. Unequal homologous recombination between invariant, TE-flanking target site duplications, is a probable mechanism to remove exclusively young TEs that are usually flanked by perfect short duplications (van de Lagemaat et al., 2005). The efficiency of such homology-dependent illegitimate recombination is also recognized in the formation of isolated-LTRs (Shirasu et al., 2000).

Finally, if the ultimate purpose of TE regulation is to achieve a condition of equilibrium in the control of their gene expression, those effects that result from excessive silencing by the host deserve to be considered as well. Although a pronounced proliferative activity of TEs is in itself deleterious, even their too efficient silencing can have equally negative consequences, particularly when the TEs associated with the host's genes are silenced. In essence, the host genome will have to avoid excessive repression of gene expression, reducing its redundancy, and with it the same evolvability, forcing the population to survive in a too narrow ecological niche. Balanced management of genetic and/or epigenetic regulation should instead allow the host to efficiently co-opt many genetic variations, which, in the long term, could prove to be evolutionarily advantageous (Schrader and Schmitz, 2019).

As in all dynamic situations in which equilibrium is a point of arrival rather than of departure, the problem arises from knowing whether the epigenetic regulation of TEs represents the functional response that the host has been able to implement to limit their proliferation or vice-versa, whether it is the very existence of epigenetic mechanisms that have allowed the TEs to spread in all eukaryotic organisms and enrich their syntactic heritage.

However, despite the apparent contradiction, it is interesting to recall the interpretation of Fedoroff (2012), according to which it was precisely the evolution of epigenetic mechanisms in prokaryotes that favoured homologous genome recombination and consequently TE accumulation. Therefore, it would not be the epigenetic mechanisms to oppose the diffusion of TEs but the opposite, namely that the TEs would have accumulated in the genomes of all living organisms because of the existence of epigenetic silencing mechanisms. Despite the plausibility of this interpretation, however, it will be necessary to undertake much more research before a shared and essentially historical interpretation can be reached regarding the origin and diffusion of TEs in eukaryotes.

6. Conclusions

In conclusion, this review has partly shown how TEs have developed strategies to ensure maximum proliferation and how the host genomes have opposed the invasion, evolving epigenetic mechanisms to silence the expression and mobility of TEs. In the absence of a functional attribution, TEs were initially interpreted as ‘selfish’ or as ‘junk.’ This is a distorted view of repeated sequences that is unfortunately still abused but no longer makes sense considering the role of TEs now recognised in the evolution of organisms. Many TEs encode a wide range of biochemical activities that normally benefit their own propagation. For example, TEs often contain promoter or enhancer elements that highjack cellular RNA polymerases for transcription, and autonomous elements encode proteins with various biochemical and enzymatic activities, all of which are necessary for the TE to replicate. Considering what has been discussed, it should also be evident that there was, and could not have been, any pre-existing TE function and that all silencing mechanisms implemented by the host genomes emerged over time, gradually transforming their relationship from competitive to cooperative.

Acknowledgements

This study was funded by the Special Fund 2019-2020 of the University of Pisa. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Declaration of competing interest

The authors declare no conflict of interests.

857 **Legend of figures.**

858 Fig. 1. Features and effects of transposable elements (TEs) into host genomes.

859 Fig. 2. Schematic representation of transposable elements (TEs) activation and
860 mobilization. The combination of TE activity and epigenetic-mediated control impact host
861 evolution.

862

References

Agius, F., Kapoor, A., Zhu, J.-K., 2006. Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. Proc. Natl. Acad. Sci. USA 103 (31), 11796-11801; https://doi.org/10.1073/pnas.0603563103.

Ali, A., Han, K., Liang, P., 2021. Role of transposable elements in gene regulation in the human genome. Life 11(2), 118. https://doi.org/10.3390/life11020118.

Alm, E., Arkin, A.P., 2003. Biological network. Curr. Opin. Struct. Biol. 13(2),193-202. https://doi.org/10.1016/s0959-440x(03)00031-9.

Almojil, D., Bourgeois, Y., Falis, M., Hariyani, I., Wilcox, J., Boissinot, S. 2021. The structural, function and evolutionary impact of transposable elements in eukaryotes. Genes 12, 918. https://doi.org/10.3390/genes.12060918.

Ayarpadikannan, S., Kim, H.-S., 2014. The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. Genomics Inform. 12, 98-104. https://doi.org/10.5808/GI.2014.12.3.98.

Aziz, M.F., Caetano-Anollés, G., 2021. Evolution of networks of protein domain organization. Sci. Rep. 11, 12075. https://doi.org/10.1038/s41598-021-90498-8.

Badouin, H., Gouzy, J., Grassa, C.J., Murat, F., Staton, S.E., Cottret, L., Lelandais-Brière, C., Owens, G.L., Carrère, S., Mayjonade, B., … Langlade, N.B., 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature 546 (7656), 148-152. https://doi.org/10.1038/nature22380

Bartlett, A.A., Hunter, R.G., 2018. Transposons, stress and the functions of the deep genome. Front. Neuroendocrin. 49, 170-174. https://doi.org/10.1016/j.yfrne.2018.03.002.

Bhattacharyya, M.K., Smith, A.M., Ellis, T.H., Hedley, C., Martin, C., 1990. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell 60(1), 115-122. https://doi.org/10.1016/0092-8674(90)90721-p.

Bhattacharyya, M.K., Martin, C., Smith, A.M., 1993. The importance of starch biosynthesis in the wrinkled seed shape character of peas studied by Mendel. Plant Mol. Biol. 22(3), 525-531. https://doi.org/10.1007/BF00015981.

Beadle, G.W., Tatum, E.L., 1941. Genetic control of biochemical reactions in *Neurospora*. Proc. Natl. Acad. Sci. USA 27(11), 499-506. https://doi.org/10.1073/pnas.27.11.499.

Belfort, M., Curcio, M.J., Lue, N.F., 2011. Telomerase and retrasposons: reverse transcriptase that shaped genomes. Proc. Natl. Acad. Sci. USA 108(51), 20304-20310. https://doi.org/10.1073/pnas.1100269109.

Bennetzen, J.L., 2005. Transposable element, gene creation and genome rearrangement in flowering plants. Curr. Opin. Genet. Dev. 15(6), 621-627. https://doi.org/10.1016/j.gde.2005.09.010.

Bennetzen, J.L., Wang, H., 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu. Rev. Plant Biol. 65(4), 505-530. https://doi.org/10.1146/annurev-arplant-050213-035811.

Biémont, C., 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186(4), 1085-1093. https://doi.org/10.1534/genetics.110.124180.

Bingham, P.M., Kidwell, M.G., Rubin, G.M., 1982. The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell 29, 995-1004.

Blumenstiel, J.P., 2019. Bith, school, work, death and resurrection: The life stages and dynamics of transposable element proliferation. Genes 10(5), 336. https://doi.org/10.3390/genes10050336.

Boeke, J.D., Devine, S.E., 1998. Yeast retrotransposons: finding a nice quiet neighborhood. Cell 93(7), 1087-1089.

Borodulina, O.R., Kramerov, D.A., 1999. Wide distribution of short interspersed elements among eukaryotic genomes. FEBS Lett. 457(3), 409-413. https://doi.org/10.1016/S0014-5793(99)01059-5.

Bourque, G., Burns, K., Gehring, M., Gorbunova, V., Seluanov, A., Hammel, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., Mager, D.L., Feschotte, C., 2018. Ten things you should know about transposable elements. Genome Biol. 19(1), 199. https://doi.org/10.1186/s13059-018-1577-z.

Bucher, E., Rinders, J., Mirouze, M., 2012. Epigenetic control of transposon transcription and mobility in *Arabidopsis*. Curr. Opin. Plant Biol. 15(5), 503-510. https://doi.org/10.1016/j.pbi.2012.08.006.

Bucheton, A., Paro, R., Sang, H.M., Pelisson, A., Finnegan, D.J., 1984. The molecular basis of I-R hybrid dysgenesis in Drosophila melanogaster: identification, cloning and properties of the I factor. Cell 38, 153-163.

929 Bundock, P., Hooykaas, P., 2005. An *Arabidopsis hAT*-like transposase is essential for
930 plant development. Nature 436(7048):282-284. https://doi.org/10.1038/nature03667.
931 Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero
932 G., Martin, C., 2012. Retrotransposons control fruit-specific, cold-dependent
933 accumulation of anthocyanins in blood oranges. Plant Cell 24(3), 1242-1255.
934 https://doi.org/10.1105/tpc.111.095232.
935 Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., 2007.
936 Discrete small RNA-generating loci as master regulators of transposon activity in
937 *Drosophila*. Cell 128 (6), 1089-1103. https://doi.org/10.1016/j.cell.2007.01.043.
938 Cain, A.K., Barquist, L., Goodman, A.L., Paulsen, I.T., Parkhill, J., van Opijenen T., 2020.
939 A decade of advances in transposon-insertion sequencing. Nat. Rev. Genet. 21(6),
940 526-540. https://doi.org/10.1038/s41576-020-0244-x.
941 Capel, J., Montero, L.M., Martinez-Zapater, J.M., Salinas, J. 1993. Non-random
942 distribution of transposable elements in the nuclear genome of plants. Nucleic Acids
943 Res. 21(10), 2369-2373. https://doi.org/10.1093/nar/21.10.2369.
944 Cariani, P., 1998. Towards an evolutionary semiotics: the emergence of new sign-
945 functions in organisms and devices, in: Van de Vijver, G., Salthe, S., Delpos, M.
946 (Eds.), Evolutionary Systems. Kluwer, Dordrecht, Holland, pp. 359-377.
947 Carter Jr., C.W., Wills, P.R., 2017. Interdependence, reflexivity, fidelity, impedance
948 matching, and the evolution of genetic coding. Mol. Biol. Evol. 35(2), 269-286.
949 https://doi.org/10.1093/molbev/msx265.
950 Casacuberta, E., 2017. *Drosophila***:** retrotransposons making up telomeres. Viruses 9(7),
951 192. https://doi.org/10.3390/v9070192.
952 Casacuberta, E., González, J., 2013. The impact of transposable elements in
953 environmental adaptation. Mol. Ecol. 22(6), 1503-1517.
954 https://doi.org/10.1111/mec.12170.
955 Cavrak, V.V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M., Mittelsten Scheid, O.,
956 2014. How a retrotransposon exploits the plant's heat stress response for its
957 activation. PLoS Genet. 10(1), e1004115.
958 https://doi.org/10.1371/journal.pgen.1004115.
959 Cerbin, S., Jiang, N., 2018. Duplication of host genes by transposable elements. Curr.
960 Opin. Genet. Dev. 49(3), 63-69. https://doi.org/10.1016/j.gde.2018.03.005.
961 Chan, S.W.L., Hendeson I.R., Jacobsen S.E., 2005. Gardening the genome DNA
962 methylation in *Arabidopsis thaliana*. Nat. Rev. Genet. 6(5), 351-360.

963 Chapman, M.A., Tang, S., Draeger, D., Nambeesan, S., Shaffer, H., Barb, J.G., Knapp, S.
964 J., Burke, J.M., 2012. Genetic analysis of floral symmetry in Van Gogh's sunflowers
965 reveals independent recruitment of *CYCLOIDEA* genes in the asteraceae. PLoS
966 Genet. 8(3), e1002628. https://doi.org/10.1371/journal.pgen.1002628.
967 Chen, J., Lu, L., Benjamin, J., Diaz, S., Hancock, C.N., Stajich, J.S., Wessler, S.R., 2019.
968 Tracking the origin of two genetic components associated with transposable element
969 bursts in domesticated rice. Nat. Commun. 10, 641. https://doi.org/10.1038/s41467-
970 019-08451-3.
971 Chen, T., Tsujimoto, N., Li, E., 2004. The PWWP domain of Dnmt3a and Dnmt3b is
972 required for directing DNA methylation to the major satellite repeats at pericentric
973 heterochromatin. Mol. Cell Biol. 24(20), 9048-9058.
974 https://doi.org/10.1128/MCB.24.20.9048-9058.2004.
975 Chénais, B., Caruso, A., Hiard, S., Casse, N., 2012. The impact of transposable elements
976 on eukaryotic genomes: From genome size increase to genetic adaptation to
977 stressful environments. Gene 509(1), 7-15.
978 https://doi.org/10.1016/j.gene.2012.07.042.
979 Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen,
980 S.E., and Fischer, R.L., 2002. DEMETER, a DNA glycosylase domain protein, is
981 required for endosperm gene imprinting and seed viability in arabidopsis. Cell 110(1),
982 33-42. https://doi.org/10.1016/S0092-8674(02)00807-3.
983 Choi, J.Y., Purugganan, M.D., 2018. Evolutionary epigenomics of retrotransposon-
984 mediated methylation spreading in rice. Mol. Biol. Evol. 35(2), 365-382.
985 https://academic.oup.com/mbe/article/35/2/365/4596560?login=true#113623882.
986 Chuong, E.B., 2013. Retroviruses facilitate the rapid evolution of the mammalian placenta.
987 Bioessays 35(10), 853–861. https://doi.org/10.1002/bies.201300059.
988 Chuong, E.B., Elde, N.C., Feschotte, C., 2017. Regulatory activities of transposable
989 elements: from conflicts to benefits. Nat. Rev. Genet. 18(11), 71-86.
990 https://doi.org/10.1038/nrg.2016.139.
991 Clark, R.M., Wagler, T.N., Quijada, P., Doebley, J., 2006. A distant upstream enhancer at
992 the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent
993 architecture. Nat. Genet. 38(5), 594-597. https://doi.org/10.1038/ng1784.
994 Cohen, I.R., 2016. Updating Darwin: Information and entropy drive the evolution of life.
995 F1000Res. 5, 2808. https://doi.org/10.12688/f1000research.10289.1.

Cohen, I.R., Harel, D., 2007. Explaining a complex living system: dynamics, multi-scaling and emergence. J. R. Soc. Interface 4(13), 175-182. https://doi.org/10.1098/rsif.2006.0173.

Cokus, S.J., Feng S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., Jacobsen, S.E., 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature 452(7184), 215-219. https://doi.org/10.1038/nature06745.

Comfort, N.C., 2001. From controlling elements to transposons: Barbara McClintock and the Nobel Prize. Trends Genet. 17(8), 475-478. https://doi.org/10.1016/S0168-9525(01)02383-6.

Cosby, R.L., Chang, N.C., Fenschotte, C., 2019. Host-transposon interaction: conflict, cooperation, and cooptation. Genes Dev. 33(17-18), 1098-1116. https://doi.org/10.1101/gad.327312.119.

Cowan, R.K., Hoen, D.R., Schoen, D.J., Bureau, T.E., 2005. *MUSTANG* is a novel family of domesticated transposase genes found in diverse angiosperms. Mol. Biol. Evol. 22(10), 2084-2089. https://doi.org/10.1093/molbev/msi202.

Czech, B., Hannon, G.J., 2016. One loop to rule them all: the Ping-Pong cycle and piRNA-guided silencing. Trends Biochem. Sci. 41(4), 324-337. https://doi.org/10.1016/j.tibs.2015.12.008.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London.

Dawkins, R., 1976. The Selfish Gene. Oxford University Press, Oxford.

Defraia, C., Slotkin, R.K., 2014. Analysis of retrotransposon activity in plants. Methods Mol. Biol. 1112, 195-210. https://doi.org/10.1007/978-1-62703-773-0_13.

de Mendoza, A., Bonnet, A., Vargas-Landin, D.B., Ji, N., Li, H., Yang, F., Li, L., Hori, K., Pflueger, J., Buckberry, S., Ohta, H., Rosic, N., Lesage, P., Lin, S., Lister, R., 2018. Recurrent acquisition of cytosine methyltransferases into eukaryotic retrotransposons. Nat. Commun. 9(1), 1341. https://doi.org/10.1038/s41467-018-03724-9.

Deniz, O., Frost, J.M., Branco, M.R., 2019. Regulation of transposable elements by DNA modifications. Nat. Rev. Genet. 20(7), 417-431. https://doi.org/10.1038/s41576-019-0106-6.

1028 Deragon, J.M., Landry, B.S., Pélissier, T., Tutois, S., Tourmente, S., Picard, G., 1994. An
1029 analysis of retroposition in plants based on a family of SINEs from *Brassica napus*. J.
1030 Mol. Evol. 39, 378-386. https://doi.org/10.1007/BF00160270.
1031 De Souza, F.S.L., Franchini, L.F., Rubinstein, M. 2013. Exaptation of transposable
1032 elements into novel *cis*-regulatory elements: is the evidence always strong?. Mol.
1033 Biol. Evol. 30(6): 1239-1251. https://doi.org/10.1093/molbev/mst045.
1034 Dewannieux, M., Esnault, C., Heidmann, T., 2003. LINE-mediated retrotransposition of
1035 marked *Alu* sequences. Nat. Genet. 35(1), 41-48. https://doi.org/10.1038/ng1223.
1036 Diao, X., Freeling, M., Lisch, D., 2006. Horizontal transfer of a plant transposon. PLoS
1037 Biol. 4(1), e5. https://doi.org/10.1371/journal.pbio.0040005.
1038 Domingues, S., da Silva, G.J., Nielsen, K.M., 2012. Integrons. Vehicles and pathways for
1039 horizontal dissemination in bacteria. Mob. Genet. Elements 2(5), 211-223.
1040 https://doi.org/10.4161/mge.22967.
1041 Doolittle, W.F., Sapienza, C., 1980. Selfish genes, the phenotype paradigm and genome
1042 evolution. Nature 284, 601-603. https://doi.org/10.1038/284601a0.
1043 Doucet, A.J., Wilusz, J.E., Miyoshi, T., Liu, Y., Moran, J.V., 2015. A 3′ poly(A) tract is
1044 required for LINE-1 retrotransposition. Mol. Cell 60(5), 728-741.
1045 https://doi.org/doi:10.1016/j.molcel.2015.10.012.
1046 Drongitis, D., Aniello, F., Fucci, L., Donizetti, A., 2019. Role of transposable elements in
1047 the different layers of gene expression regulation. Int. J. Mol. Sci. 20(22), 5755.
1048 https://doi.org/10.3390/ijms20225755.
1049 Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A.,
1050 Terragni, J., Chin, H.G., Tu, A., Hetzel, J., Wohlschlegel, J.A., Pradhan, S., Patel, D.
1051 J., Jacobsen, S.E., 2012. Dual binding of chromomethylase domains to H3K9me2-
1052 containing nucleosomes directs DNA methylation in plants. Cell 151(1), 167-180.
1053 https://doi.org/10.1016/j.cell.2012.07.034
1054 Dubin, M.J., Mittelsten Scheid, O., Becker C., 2018. Transposons: a blessing curse. Curr.
1055 Opin. Plant Biol. 42(4), 23-29. https://doi.org/10.1016/j.pbi.2018.01.003.
1056 Eickbush, T.H., Jamburuthugoda, V.K., 2008. The diversity of retrotransposons and the
1057 properties of their reverse transcriptases. Virus Res. 134(1-2), 221-234.
1058 https://doi.org/10.1016/j.virusres.2007.12.010.
1059 Eldredge, N., Gould, S.J., 1972. Punctuate equilibria: An alternative to phyletic gradualism,
1060 in: Schopf, T.J.M. (Ed.), Model in Paleobiology. Freeman, Cooper, and Company,
1061 San Francisco, CA, pp. 82-115.

Fambrini, M., Salvini, M., Pugliesi, C., 2011. A transposon-mediate inactivation of a *CYCLOIDEA*-like gene originates polysymmetric and androgynous ray flowers in *Helianthus annuus*. Genetica 139(12), 1521-1529. https://doi.org/10.1007/s10709-012-9652-y.

Fambrini, M., Basile, A., Salvini, M., Pugliesi, C., 2014a. Excisions of a defective transposable CACTA element (*Tetu1*) generate new alleles of a *CYCLOIDEA*-like gene of *Helianthus annuus*. Gene 549(1), 198-207. https://doi.org/10.1016/j.gene.2014.07.018.

Fambrini, M., Salvini, M., Basile, A., Pugliesi, C., 2014b. Transposon-dependent induction of Vincent van Gogh's sunflowers: exceptions revealed. genesis 52(4), 315-327. https://doi.org/10.1002/dvg.22743.

Fambrini, M., Bellanca, M., Costa Muñoz, M., Usai, G., Cavallini, A., Pugliesi, C., 2018. Ligulate inflorescence of *Helianthus* × *multiflorus*, cv. Soleil d'Or, correlates with a mis-regulation of a *CYCLOIDEA* gene characterised by insertion of a transposable element. Plant Biol. 20(6), 956-967. https://doi.org/10.1111/plb.12876.

Fambrini, M., Usai, G., Vangelisti, A., Mascagni, F., Pugliesi, C., 2020. The plastic genome: The impact of transposable elements on gene functionality and genomic structural variations. genesis 58(12), e23399. https://doi.org/10.1002/dvg.23399.

Feng, S., Jacobsen, S.E., Reik, W., 2010. Epigenetic reprogramming in plant and animal development. Science 330(6004), 622-627. https://doi.org/10.1126/science.1190614.

Fedoroff, N., Schläppi, M., Raina, R., 1995. Epigenetic regulation of the maize *Spm* transposon. BioEssays, 17(4), 291-297. https://doi.org/10.1002/bies.950170405.

Fedoroff, N.V., Wessler, S., Shure, M., 1983. Isolation of the transposable maize controlling element Ac and Ds. Cell 35(1), 235-242. https://doi.org/10.1016/0092-8674(83)90226-x.

Fedoroff, N.V., 2012.Transposable elements, epigenetics, and genome evolution. Science 338(6108), 758-767. https://doi.org/10.1126/science.338.6108.758.

Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. 9(6), 397-405. https://doi.org/10.1038/nrg2337.

Feschotte, C., Pritham, E.J., 2007. DNA-transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41(12), 331-368. https://doi.org/10.1146/annurev.genet.40.110405.090448.

Finnegan, D.J., 2012. Retrotransposon. Curr. Biol. 22(11), R-432-R437. https://doi.org/10.1016/j.cub.2012.04.025.

1096 Fulcher, N., Teubenbacher, A., Kerdaffrec, E., Farlow, A., Nordborg, M., Riha, K., 2014.
1097 Genetic architecture of natural variation of telomere length in *Arabidopsis thaliana*.
1098 Genetics 199(2), 625-635. https://doi.org/10.1534/genetics.114.172163.
1099 Gao, C., Ren, X., Mason, A.S., Liu, H., Xiao, M., Li, J., Fu, D., 2014. Horizontal gene
1100 transfer in plants. Funct. Integ. Genom. 14(1), 23-29. https://doi.org/10.1007/s10142-
1101 013-0345-0.
1102 Gao, X., Hou, Y., Ebina, H., Levin, H.L., Voytas D.F., 2008. Chromodomains direct
1103 integration of retrotransposons to heterochromatin. Genome Res. 18, 359-369
1104 https://doi.org/10.1101/gr.7146408.
1105 Garavís, M., González, C., Villasante, A., 2013. On the origin of the eukaryotic
1106 chromosome: The role of noncanonical DNA structure in telomere evolution. Genome
1107 Biol. Evol. 5(6), 1142-1150. https://doi.org/10.1093/gbe/evt079.
1108 Garcia-Perez, J.L., Widmann, T.J., Adams, I.R., 2016. The impact of transposable
1109 elements on mammalian development. Development 143(22), 4101-4114.
1110 https://doi.org/10.1242/dev.132639.
1111 Gehring, M., Bubb, K.L., Henikoff, S., 2009. Extensive demethylation of repetitive elements
1112 during seed development underlies gene imprinting. Science 324(5933), 1447-1451.
1113 https://doi.org/10.1126/science.1171609.
1114 Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V., Martienssen, R.A., 2002. Dependence
1115 of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene *DDM1*.
1116 Science 297(5588), 1871-1873. https://doi.org/10.1126/science.1074950.
1117 Gentlemen, J.F., Mullin, R.C. 1989. The distribution of the frequency of occurrence of
1118 nucleotide subsequences, based on their overlap capability. Biometrics 45(1), 35-52.
1119 https://doi.org/10.2307/2532033.
1120 Ghildiyal, M., Zamore, P.D., 2009. Small silencing RNAs: an expanding universe. Nat Rev
1121 Genet. 10 (2), 94-108. https://doi.org/10.1038/nrg2504.
1122 Gierl, A., Saedler H., Peterson, P.A., 1989. Maize transposable elements. Annu. Rev.
1123 Genet. 23(12), 71-85. https://doi.org/10.1146/annurev.ge.23.120189.000443.
1124 Gilbert, C., Feschotte, C., 2018. Horizontal acquisition of transposable elements and viral
1125 sequences: patterns and consequences. Curr. Opin. Genet. Dev. 49(4), 15-24.
1126 https://doi.org/10.1016/j.gde.2018.02.007.
1127 Gilbert, S.F., Sarkar, S., 2000. Embracing complexity: organicism for the 21st century.
1128 Dev. Dyn. 219(1), 1-9. https://doi.org/10.1002/1097-0177(2000)9999:9999<::AID-
1129 DVDY1036>3.0.CO;2-A.

Gill, R.A., Scossa, F., King, G.J., Golicz, A., Tong, C., Snowdon, R.J., Fernie, A.R., 2021. On the role of transposable elements in the regulation of gene expression and subgenomic interaction in crop genomes. Crit. Rev. Plant Sci 40(2), 157-189. https://doi.org/10.1080/07352689.2021.1920731.

Giordani, T., Cavallini, A., Natali, L., 2014. The repetitive component of the sunflower genome. Curr. Plant Biol. 1(8), 45-54. https://doi.org/10.1016/j.cpb.2014.05.001.

Gladyshev, E.A., Arkhipova, I.R., 2007. Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. Proc. Natl. Acad. Sci. USA 104(22), 9352-9357. https://doi.org/10.1073/pnas.0702741104.

Gong, Z., Morales-Ruiz, T., Ariza, R.R., Rold_an-Arjona, T., David, L., and Zhu, J.K., 2002. *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNAglycosylase/lyase. Cell 111(6), 803-814. https://doi.org/10.1016/s0092-8674(02)01133-9.

Gould, S.J, Eldredge, N. 1993. Puncuated equilibrium comes from ages. Nature 366(6452), 223-227. https://doi.org/10.1038/366223a0.

Grandbastien, M.A., 1998. Activation of plant retrotransposons under stress conditions. Trends Plant Sci. 3(5), 181-187. https://doi.org/10.1016/S1360-1385(98)01232-1.

Greene, B., Walko, R., Hake, S., 1994. Mutator insertions in an intron of the maize *Knotted1* gene result in dominant suppressible mutations. Genetics 138(4), 1275-1285.

Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G., Kapraun, D.F., Greilhuber, J., Bennet, M.D., 2007. Eukaryotic genome size databases. Nucleic Acids Res. 35(Database issue), D332-D338. https://doi.org/10.1093/nar/gkl828.

Guio, L., González, J., 2019. New insights on the evolution of genome content: population dynamics of transposable elements in flies and humans. Methods Mol. Biol.1910, 505-530. https://doi.org/10.1007/978-1-4939-9074-0_16.

Habibi, L., Pedran, M., AmirPhirozy, A., Bonyadi, K., 2015. Mobile DNA elements: The seeds of organic complexity on Earth. DNA Cell Biol., 34(10), 597-609. http://doi.org/10.1089/dna.2015.2938.

Henderson, R., Deleris, A., Wong, W., Zhong, X., Chin, H.G., Horwitz, G.A. Kelly, K.A., Pradhan, S., Jacobsen, S.E., 2010. The *De Novo* cytosine methyltransferase DRM2 requires Intact UBA domains and a catalytically mutated paralog DRM3 during RNA–

directed DNA methylation in *Arabidopsis thaliana.* PLoS Genet. 6(10), e1001182. https://doi.org/10.1371/journal.pgen.1001182.

Hershberger, R.J., Benito, M.I., Hardeman, K.J., Warren, C., Chandler, V.L., Walbot, V. 1995. Characterization of the major transcripts encoded by the regulatory MuDR transposable element of maize. Genetics 140(3), 1087-1098.

Hiraizumi, Y., 1971. Spontaneous recombination in *Drosophila melanogaster* males. Proc. Natl. Acad. Sci. USA. 68(2), 268-270.

Hirsh, C.D., Springer, N.M., 2017. Transposable element influences on gene expression in plants. Biochim. Biophys. Acta Gene Regul. Mech. 1860(1), 157-165. https://doi.org/10.1016/j.bbagrm.2016.05.010.

Hoen, D.R., Bureau, T.E., 2015. Discovery of novel genes derived from transposable elements using integrative genomic analysis. Mol. Biol. Evol. 32(6), 1487-1506. https://doi.org/10.1093/molbev/msv042.

Hoffmeyer, J., Emmeche, C., 1991. Code-duality and the semiotics of nature. In: Anderson, M., Merrell F. (Eds.), On Semiotic Modeling. Mouton de Gruyter, Berlin, New York, pp. 117-166.

Hoffmeyer J., 2008. Semiotic scaffolding of living systems. In: Barbieri M. (Ed.) Introduction to Biosemiotics. Springer, Dordrecht. https://doi.org/10.1007/1-4020-4814-9_6.

Hollister, J.D., Gaut, B.S., 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 19(8), 1419-1428. https://doi.org/10.1101/gr.091678.109.

Hordijk, W., Steel, M., 2015. Autocatalytic sets and boundaries. J. Systems Chem. 6(1), 1. https://doi.org/10.1186/s13322-014-0006-2.

Hordijk, W., Steel, M., 2018. Autocatalytic networks at the basis of life's origin and organization. Life 8, 62. https://doi.org/10.3390/life8040062.

Hosaka, A., Kakutami, T., 2018. Transposable elements, genome evolution and transgenerational epigenetic variation. Curr. Opin. Genet. Dev. 49(3), 43-48. https://doi.org/10.1016/j.gde.2018.02.012.

Huang, C.R.L., Burns, K.H., Bok, J.D., 2012. Active transposition in genomes. Annu. Rev. Genet. 46, 651-675. https://doi.org/10.1146/annurev-genet-110711-155616.

Hudson, M.E., Lisch, D.R., Quail, P.H., 2003. The *FHY3* and *FAR1* genes encode transposaserelated proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. Plant J. 34(4), 453-471. https://doi.org/10.1046/j.1365-313X.2003.01741.x.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., Paszkowski, J., 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472(7341), 115-119. https://doi.org/10.1038/nature09861.

Ito, H., 2013. Small RNAs and regulation of transposons in plants. Genes Genet. Syst. 88(1), 3-7. https://doi.org/10.1266/ggs.88.3.

Ivancevic, A.M., Kortschak, R.D., Bertozzi, T., Adelson, D.L., 2016. LINEs between species: evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. Genome Biol. Evol. 8(11), 3301-3322. https://doi.org/10.1093/gbe/evw243.

Ivancevic, A.M., Kortschak, R.D., Bertozzi, T., Adelson, D.L., 2018. Horizontal transfer of BovB and L1 retrotransposon in eukaryotes. Genome Biol. 19(1), 85. https://doi.org/10.1186/s13059-018-1456-7.

Jacob, F., 1976. The Logic of Life. Vanguard: N.Y. (Orig. published in French, Editions Gallimard, 1970).

Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol., 3(3), 318-356. https://doi.org/10.1016/S0022-2836(61)80072-7.

Janga, S.C., Salgado, H., Martínez-Antonio, A., 2009. Transcriptional regulation shapes the organization of genes on bacterial chromosomes. Nucleic Acids Res. 37(11), 3680–3688. https://doi.org/10.1093/nar/gkp231.

Jangam, D., Feschotte, C., Betrán, E., 2017. Transposable element domestication as an adaptation to evolutionary conflicts. Trends Genet. 33(11), 817-831. https://doi.org/10.1016/j.tig.2017.07.011.

Jansz, N., 2019. DNA methylation dynamics at transposable elements in mammals. Essays Biochem. 63(6), 677-689. https://doi.org/10.1042/EBC20190039.

Jiang, N., Jordan, I.K., Wessler, S.R., 2002. *Dasheng* and *RIRE2*. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. Plant Physiol. 130(4), 1697-1705. https://doi.org/10.1104/pp.015412.

Jiang, N., Feschotte, C., Zhang, X., Wessler, S.R., 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). Curr. Opin. Plant Biol. 7, 115-119. https://doi.org/10.1016/j.pbi.2004.01.004.

Jiang, S.Y., Ramachandran, S., 2013. Genome-wide survey and comparative analysis of LTR retrotransposons and their captured genes in rice and sorghum. PLoS One 8(7), e71118. https://doi.org/10.1371/journal.pone.0071118.

Joly-Lopez, Z., Bureau, T.E., 2018. Exaptation of transposable element coding sequences. Curr. Opin. Genet. Dev. 49, 34-42. https://doi.org/10.1016/j.gde.2018.02.011.

Kashkush, K., Feldman, M., Levy, A.A., 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat. Genet. 33(1), 102-106. https://doi.org/10.1038/ng1063.

Kashkush, K., Khasdan, V., 2007. Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. Genetics 177(4), 1975-1985. https://doi.org/10.1534/genetics.107.080234.

Kauffman, S.A., 1986. Autocatalytic sets of proteins. J. Theor. Biol. 119(1), 1-24. https://doi.org/10.1016/s0022-5193(86)80047-9.

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J.J., Koornneef, M., Kakutani, T., 2007. Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. Plant J. 49(1), 38-45. https://doi.org/10.1111/j.1365-313X.2006.02936.x.

Klose, R.J., Yamane, K., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., Zhang, Y., 2006. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature 442(7100), 312-316. https://doi.org/10.1038/nature04853.

Koonin, E.V., 2006. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol. Direct 1, 22. https://doi.org/10.1186/1745-6150-1-22.

Koonin, E.V., Makarova, K.S., Wolf, Y.I., Krupovic, M., 2020. Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. Nat. Rev. Genet. 21(10), 119-131. https://doi.org/10.1038/s41576-019-0172-9.

Kordyukova, M., Olovnikov, I., Kalmykova, A., 2018. Transposon control mechanism in telomere biology. Curr. Opin. Genet. Dev. 49(3), 56-62. https://doi.org/10.1016/j.gde.2018.03.002.

Kremer, C.S., Linquist, S., Saylor, B., Elliott, T.A., Gregory, T.R., Cottenie, K., 2020. Transposable element persistence via potential genome-level ecosystem engineering. BMC Genom. 21, 367. https://doi.org/10.1186/s12864-020-6763-1.

Kuhn, T., 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago, IL, USA.

Kumar, A., Bennetzen, J.L., 1999. Plant retrotransposons. Annu. Rev. Genet. 33(12), 479-532. https://doi.org/10.1146/annurev.genet.33.1.479.

Kumar, S., Cheng, X., Pflugrath, J.W., Roberts, R.J., 1992. Purification, crystallization, and preliminary X-ray diffraction analysis of an M. Hhal-AdoMet complex. Biochemistry 31(36), 8648-8653. https://doi.org/10.1021/bi00151a035.

Labrador, M., Corces, V.G., 1997. Transposable element-host interaction: regulation of insertion and excision. Annu. Rev. Genet. 31, 381-404. https://doi.org/10.1146/annurev.genet.31.1.381.

Lambowitz, A.M., Zimmerly, S., 2004. Mobile group II introns. Annu. Rev. Genet. 38, 1-35. 10.1146/annurev.genet.38.072902.091600.

Lambowitz, A.M., Zimmerly, S., 2011. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb. Perspect. Biol. 3(8), a003616. https://doi.org/10.1101/cshperspect.a003616.

Lang-Mladek, C., Popova, O., Kiok, K., Berlinger, M., Rakic, B., Aufsatz, W., Jonak, C., Hauser, M.-T., Luschnig, C., 2010. Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in *Arabidopsis*. Mol. Plant. 3(3), 594-602. https://doi.org/10.1093/mp/ssq014.

Law, J.A., Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11(2), 204-220. https://doi.org/10.1038/nrg2719.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., Carthew, R.W., 2004. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell 117(1), 69-81. https://doi.org/10.1016/s0092-8674(04)00261-2.

Lee, S.-I., Kim, N.-S., 2014. Transposable elements and genome size variations in plants. Genomics Inform. 12(3), 87-97. https://doi.org/10.5808/GI.2014.12.3.87.

Leuschner, P.J., Ameres, S.L., Kueng, S., Martinez, J., 2006. Cleavage of the siRNA passenger strand during RISC assembly in human cells. EMBO Rep. 7(3), 314-320. https://doi.org/10.1038/sj.embor.7400637.

Li, S., Vandivier, L.E., Tu, B., Gao, L., Won, S.Y., Li, S., Zheng, B., Gregory, B.D., Chen, X., 2015. Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in *Arabidopsis* reveals features and regulation of siRNA biogenesis. Genome Res. 25, 235-245. https://doi.org/10.1101/gr.182238.114.

Lippman, Z., May, B., Yordan, C., Singer, T., Martienssen, R., 2003. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and

histone modification. PLoS Biol. 1(3), e67. https://doi.org/10.1371/journal.pbio.0000067.

Lippman, Z.B., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B.P., Kasschau, K.D. Carrington, J.C., Doerge, R.W., Colot, V., Martienssen, R., 2004. Role of transposable elements in heterochromatin and epigenetic control. Nature 430(6998), 471-476. https://doi.org/10.1038/nature02651.

Lisch, D., 2013. How important are transposons for plant evolution? Nat. Rev. Genet. 14(1), 49-61. https://doi.org/10.1038/nrg3374.

Liu, J., He, Y., Amasino, R., Chen, X., 2004. siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. Genes Dev. 18(23), 2873-2878. https://doi.org/10.1101/gad.1217304.

Llorens, C., Futami, R., Covelli, L., Domínguez-Escribá, L., Viu, J.M., Tamarit, D., Aguilar-Rodríguez, J., Vicente-Ripolles, M., Fuster, G., Bernet, G.P., Maumas, F., Munoz-Pomer, A., Sempere, J.M. Latorre, A., Moya, A., 2011. The *Gypsy* Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. 39, (Data Base Issue), D70-D74. https://doi.org/10.1093/nar/gkq1061.

Luo, S., Lu, J., 2017. Silencing of transponable elements by piRNAs in *Drosophila*: An evolutionary perpective. Genom. Proteom. Bioinf. 15()3, 164-176. https://doi.org/10.1016/j.gpb.2017.01.006.

Luo, S., Zhang, H., Duan, Y., Yao, X., Clark, A.G., Lu, J., 2020. The evolutionary arms race between transposable elements and piRNAs in *Drosophila melanogaster*. BMC Evol. Biol. 20(1), 14. https://doi.org/10.1186/s12862-020-1580-3.

Lv, H., Dao, F.Y., Zhang, D., Yang, H., Lin, H., 2021. Advances in mapping the epigenetic modifications of 5-methylcytosine (5mC), N6-methyladenine (6mA), and N4-methylcytosine (4mC). Biotechnol. Bioeng. https://doi.org/10.1002/bit.27911.

Martin, S.L., 2018. Retrotransposons: On the move. eLife 7, e30058. https://doi.org/10.7554/eLife.34901.

Majumdar, S., Rio, D.C., 2015. P transposable elements in Drosophila and other eukaryotic organisms. Microbiol. Spectr. 3(2), MDNA3-0004. https://doi.org/10.1128/microbiolspec.MDNA3-0004-2014.

Makarevitch, I., Waters, A.J., West, P.T., Stitzer, M., Hirsch, C.N., Ross-Ibarra, J., Springer, N.M., 2015. Transposable elements contribute to activation of maize genes

in response to abiotic stress. PloS Genet. 11(1), e1004915. https://doi.org/10.1371/journal.pgen.1004915.

Mascagni, F., Barghini, E., Giordani, T., Rieseberg, L.H., Cavallini, A., Natali, L., 2015. Repetitive DNA and plant domestication: variation in copy number and proximity to genes of LTR-retrotransposons among wild and cultivated sunflower (*Helianthus annuus*) genotypes. Genome Biol. Evol. 7(12), 3368-3382. https://doi.org/10.1093/gbe/evv230.

Mascagni, F., Giordani, T., Ceccarelli, M., Cavallini, A., Natali, L., 2017. Genome-wide analysis of LTR-retrotransposon diversity and its impact on the evolution of the genus *Helianthus* (L). BMC Genomics 18(1), 634. https://doi.org/0.1186/s12864-017-4050-6.

McClintock, B., 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA 36(6), 344-355. https://doi.org/10.1073/pnas.36.6.344.

McClintock, B., 1953. Induction of instability at selected loci in maize. Genetics 38(6), 579-599.

McClintock, B., 1956. Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology,* Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 197-216. https://doi.org/10.1101/sqb.1956.021.01.017

McClintock, B., 1961. Some parallels between gene control systems in maize and in bacteria. Am. Nat. 95(884), 265-277.

McClintock, B., 1984. The significance of responses of the genome to challenge. Science 226(4276), 792-801. https://doi.org/10.1126/science.15739260.

Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H., Kakutani, T., 2009. An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. EMBO J. 28(8), 1078-1086. https://doi.org/10.1038/emboj.2009.59.

Monod, J., 1970. Le hasard et la necessité: essai sur la philosophie naturelle de la biologie moderne. Editions du Seuil, Paris.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., Rafalski, A., 2005. Gene duplication and exon shuffling by *helitron*-like transposons generate intraspecies diversity in maize. Nat. Genet. 37(9), 997-1002. https://doi.org/10.1038/ng1615.

Muñoz-López, M., García-Pérez, J.L., 2010. DNA transposons: nature and applications in genomics. Curr. Genom. 11(2), 115-128. https://doi.org/10.2174/138920210790886871.

Muskhelishvili, G., Travers, A., 2013, Integration of syntactic and semantic properties of the DNA code reveals chromosomes as thermodynamic machines converting energy into information. Cell. Mol. Life Sci. 70, 4555–4567. https://doi.org/10.1007/s00018-013-1394-1.

Muyle, A., Deymour, D., Darzentas, N., Primetis, E., Gaut, B.S., Bousios, A., 2021. Gene capture by transposable elements leads to epigenetic conflict in maize. Mol. Plant 14(2), 237-252. https://doi.org/10.1016/j.molp.2020.11.003.

Nekrutenko, A., Li, W. H., 2001. Transposable elements are found in a large number of human protein-coding genes. Trends Genet. 17(11), 619-621. https://doi.org/10.1016/s0168-9525(01)02445-3.

Neumann, P., Novák, P., Hoštáková, N., Macas, J., 2019. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mobile DNA 10, 1. https://doi.org/10.1186/s13100-018-0144-1.

Nicetto, D., Zaret, K.S., 2019. Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr. Opin. Genet. Dev. 55, 1-10. https://doi.org/10.1016/j.gde.2019.04.013.

Nishihara, H., 2020. Transposable elements as genetic accelerators of evolution: contribution to genome size, gene regulatory network rewiring and morphological innovation. Genes Genet. Syst. 94(6), 269-281. https://doi.org/10.1266/ggs.19-00029.

Nishiyama, E., Ohshima, K., 2018. Cross-kingdom commonality of a novel insertion signature of RTE-related short retroposons. Genome Biol. Evol. 10(6), 1471-1483. https://doi.org/10.1093/gbe/evy098.

Oliver, K.R., Greene, W.K., 2009. Transposable elements: powerful facilitators of evolution. BioEssays 31(7), 703-714. https://doi.org/10.1002/bies.200800219.

Oliver, K.R., McComb, J.A., Greene, W.K., 2013. Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol. Evol. 5(10), 1886-1901. https://doi.org/10.1093/gbe/evt141.

Orgel, L.E., Crick, F.H., 1980. Selfish DNA: the ultimate parasite. Nature 284(5757), 604-607. https://doi.org/10.1038/284604a0.

Orozco-Arias, S., Isaza, G., Guyot, R., 2019. Retrotransposons in plant genomes: structure, identification, and classification through bioinformatics and machine learning. Int. J. Mol. Sci. 20(15), 3837. https://doi.org10.3390/ijms20153837.

Pace, J.K., Gilbert, C., Clark, M.S., Feschotte, C., 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc. Natl. Acad. Sci. USA, 105(44), 17023-17028. https://doi.org/10.1073/pnas.0806548105.

Pardue, M.-L., DeBaryshe, P.G., 2011. Adapting to life at the end of the Line. How *Drosophila* telomeric retrotransposons cope with their job. Mob. Genet. Elements 1(2), 128-134. https://doi.org/10.4161/mge.1.2.16914.

Pattee, H., 1972. Laws and constraints, symbols and languages, in: Waddington, C.H., (Ed.), Towards a Theoretical Biology, Vol. 4. University of Edinburgh Press, Edinburgh, UK, pp: 248-258.

Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N., Mittelsten Scheid, O., 2010. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in *Arabidopsis*. Plant Cell 22(9), 3118-3129. https://doi.org/10.1105/tpc.110.078493.

Pradhan, S., Adams, R.L.P., 1995. Distinct CG and CNG DNA methyl-transferases in *Pisum sativum.* Plant J. 7(3), 471-81. 10.1046/j.1365-313x.1995.7030471.x.

Rayan, N.A., Del Rosario, R.C.H., Prabhakar, S., 2016. Massive contribution of transposable elements to mammalian regulatory sequences. Semin. Cell Dev. Biol. 57, 51-56. https://doi.org/10.1016/j.semcdb.2016.05.004.

Rech, G.E., Bogaerts-Márquez, M., Barrón, M.G., Merenciano, M., Villanueva-Cañas, J.L., Horváth, V., Fiston-Lavier, A.S., Luyten, I., Venkataram, S., Quesneville, H., Petrov, D.A., González, J., 2019. Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. PLoS Genet. 15(2), e1007900. https://doi.org/10.1371/journal.pgen.1007900.

Richardson, S.R., Gerdes, P., Gerhardt, D.J., Sanchez-Luque, F.J., Bodea, G.O., Munoz-Lopez, M., Jesuadian, J.S., Kempen, M.-J.H.C., Carreira, P.E., Jeddeloh, J.A., Garcia-Perez, J.L., Kazazian Jr., H.H., Ewing, A.D., Faulkner, G.J., 2017. Heritable L1 retrotransposition in the mouse primordial germline and early embryo. Genome Res. 27(8), 1395-1405. https://doi.org/10.1101/gr.219022.116.

Rigal, M., Mathieu, O., 2011. A ‘‘Mille-Feuille’’ of silencing: epigenetic control of transposable elements. Biochem. Biophys. Acta 1809(8), 452-458. https://doi.org/10.1016/j.bbagrm.2011.04.001.

Rigal, M., Becker, C., Pélissier, T., Pogorelcnik, R., Devos, J., Ikeda, Y., Weigel, D., Mathieu, O., 2016. Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in *Arabidopsis thaliana* F_1 epihybrids.

Proc. Natl. Acad. Sci. USA 113(14), E2083-E2092. https://doi.org/10.1073/pnas.1600672113.

Rhoades, M.M., 1936. A cytogenetic study of a chromosome fragment in maize. Genetics 21(5), 491-502.

Roccaro, M., Li, Y., Masiero, S., Saedler, H., Sommer, H., 2005. ROSINA (RSI), a novel protein with DNA-binding capacity, acts during floral organ development in *Antirrhinum majus*. Plant J. 43(2), 238-250. https://doi.org/10.1111/j.1365-313X.2005.02446.x.

Roccaro, M., Li, Y., Sommer, H., Saedler, H., 2007. ROSINA (RSI) is part of a CACTA transposable element, *TamRSI*, and links flower development to transposon activity. Mol. Genet. Genom. 278(6), 243-254. https://doi.org/10.1007/s00438-007-0245-x.

Rodon Fores, J., Criado-Gonzales, M., Chaumont, A., Carvalho, A., Blanck, C., Schmutz, M., Boulmedais, F. Schaaf, P., Jierry, L., 2020. Autonomous growth of a spatially localized supramolecular hydrogel with autocatalytic ability. Angewandte Chemie 59(34), 14558-14563. https://doi.org/10.1002/anie.202005377.

Rodriguez-Terrones, D., Tottes-Padilla, M.E., 2018. Nimble and ready to mingle: Transposon outbursts of early development. Trends Genet. 34(10), 806-820. https://doi.10.1016/j.tig.2018.06.006.

Roll-Hansen, N., 2014. The holist tradition in twentieth century genetics. Wilhelm Johannsen's genotype concept. J. Physiol. 592(11), 2431-2438. https://doi.10.1113/jphysiol.2014.272120.

Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., Giraut, L., Després, B., Drevensek, S., Barneche, F., Dèrozier, S., Brunaud, V., Aubourg, S., Schnittger, A., Bowler, C., Martin-Magniette, M.L., Robin, S., Caboche, M., Colot, V., 2011. Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. EMBO J. 30(10), 1928-1938. https://doi.org/10.1038/emboj.2011.103.

Sabot, F., Schulman, A.H., 2006. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity 97(6), 381-388. https://doi.org/10.1038/sj.hdy.6800903.

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., Bennetzen, J.L., 1998. The paleontology of intergene retrotransposons of maize. Nat. Genet. 20(1), 43-45. https://doi.org/10.1038/1695.

Sassaman, D.M., Dombroski, B.A, Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J, Gabriel, A., Swergold, G.D., Kazazian Jr., H.H., 1997. Many human L1 elements are capable of retrotransposition. Nat. Genet. 16(1), 37-43. https://doi.org/10.1038/ng0597-37.

Saze, H., Kakutani, T., 2007. Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. EMBO J. 26(15), 3641-3652. https://doi.org/10.1038/sj.emboj.7601788.

Saze, H., Shiraishi, A., Miura, A., Kakutani, T., 2008. Control of genic DNA methylation by a jmjC domain-containing protein in *Arabidopsis thaliana*. Science 319(5862), 462-465. https://doi.org/10.1126/science.1150987.

Saze, H., Tsugane, K., Kanno, T., Nishimura, T., 2012. DNA methylation in plants: relationship to small RNAs and histone modifications, and functions in transposon inactivation. Plant Cell Physiol. 53(5), 766-784. https://doi.org/10.1093/pcp/pcs008.

Saze, H., Kitayama, J., Takashima, K., Miura, S., Harukawa, Y., Ito, T., Kakutani, T., 2013. Mechanism for full-length RNA processing of *Arabidopsis* genes containing intragenic heterochromatin. Nat. Commun. 4, 2301. https://doi.org/10.1038/ncomms3301.

Schmid, C., Maraia, R., 1992. Transcriptional regulation and transpositional selection of active SINE sequences. Curr. Opin. Genet. Dev. 2(6), 874-882. https://doi.org/10.1016/S0959-437X(05)80110-8.

Schrader, T., Schmitz, J., 2019. The impact of transposable elements in adaptive evolution. Mol. Ecol. 28(6), 1537-1549. https://doi.org/10.1111/mec.14794.

Selinger, D.A., Chandler, V.L., 2001. *B-Bolivia*, an allele of the maize *b1* gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream. Plant Physiol. 125(3), 1363-1379. https://doi.org/10.1104/pp.125.3.1363.

Servant, G., Deininger, P.L., 2015. Insertion of retrotransposons at chromosome ends: adaptive response to chromosome maintenance. Front. Genet. 6, 358. https://doi.org/10.3389/fgene.2015.00358.

Shannon, C.E., Weaver, W., 1964. The Mathematical Theory of Communication. The University of Illinois, Urbana, USA.

Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. BioEssays, 22(2), 148-160. https://doi.org/10.1002/(SICI)1521-1878(200002)22:2<148::AID-BIES6>3.0.CO;2-Z.

Shen, Y., Conde Silva, N., Audonnet, L., Servet, C., Wei, W., Zhou, D.-X., 2014. Over-expression of histone H3K4 demethylase gene *JMJ15* enhances salt tolerance in *Arabidopsis*. Front. Plant Sci., 5, 290. https://doi.org/10.3389/fpls.2014.00290.

Shirasu, K., Schulman, A. H., Lahaye, T., Schulze-Lefert, P., 2000. Acontiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res. 10(7), 908-915. https://doi. org/10.1101/gr.10.7.908.

Singh, R.M., 2015. Limits of imagination: the 150th Anniversary of Mende's Laws and why Mendel failed to see the importance of his discovery for Darwin's theory of evolution. Genome 58(9), 415-421. https://doi.org/10.1139/gen-2015-0107.

Slotkin, R.K., Martienssen, R.A., 2007. Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8(4), 272-285. https://doi.org/10.1038/nrg2072.

Slotkin, R.K., Vaughn, M.W., Borges, F., Tanurdzić, M., Becker, J.D., Feijó, J.A., Martienssen, R.A., 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136(3), 461-472. https://doi.org/10.1016/j.cell.2008.12.038.

Song, M.J., Schaack, S., 2018. Evolution conflict between mobile DNA and host genome. Am. Nat. 192(2), 263-273. https://doi.org/10.1086/698482.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., Peeters, A., 2000. The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol. Cell 6(4), 791-802. https://doi.org/10.1016/s1097-2765(05)00090-0.

Sotero-Caio, C.G., Platt II, R.N., Suh, A., Ray, D.A., 2017. Evolution and diversity of transposable elements in vertebrate genomes. Gen. Biol. Evol. 9(1), 161-177. https://doi.org/10.1093/gbe/evw264.

Stoye, J.P., 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat. Rev. Microbiol. 10(6), 395-406. https://doi.org/10.1038/nrmicro2783.

Strogatz, S. H., 2001. Exploring complex networks. Nature, 410(6825), 268–276. https://doi.org/10.1038/35065725.

Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., Jacobsen, S.E., 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. Cell 152(1-2), 352-364. https://doi.org/10.1016/j.cell.2012.10.054.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., Jacobsen, S.E., 2014. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. Nat. Struct. Mol. Biol. 21(1), 64-72. https://doi.org/10.1038/nsmb.2735.

Stuart, T., Eichten, S.R., Cahn, J., Karpievitich, Y.V., Borevitz, J.O., Lister, R., 2016. Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. eLife 5, 20777. https://doi.org/10.7554/eLife.20777.

Studer, A., Zhao, Q., Ross-Ibarra, J., Doebley, J., 2011. Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat. Genet. 43(11), 1160-1163. https://doi.org/10.1038/ng.942.

Sun, Q., Zhou, D.-X., 2008. Rice jmjC domain-containing gene *JMJ706* encodes H3K9 demethylase required for floral organ development. Proc. Natl. Acad. Sci. USA 105(36), 13679-13684. https://doi.org/10.1073/pnas.0805901105.

Sun, Y.-W., Liu, Y.-Y., Wu, H., Wang, L.-F., Liu, J.-H., Yuan, L., Pan, Y.-S., He, D.-D., Hu, G.-Z., 2019. IS26-flanked composite transposon Tn6539 carrying the *tet*(M) gene in IncH12-type conjugative plasmids from *Escherichia coli* isolated from ducks in China. Front. Microbiol. 9, 3168. https://doi.org/10.3389/fmicb.2018.03168.

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M.P., Wang, T., 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res. 24(12), 1963-1976. https://doi.org/10.1101/gr.168872.113.

Suzuki, M.M., Bird, A., 2008. DNA methylation landaskapes: provocative insights from epigenomics. Nat. Rev. Genet. 9(6), 465-476. https://doi.org/10.1038/nrg2341.

Tavakoli N.P., Derbyshire, K.M., 2001. Tipping the balance between replicative and simple transposition. EMBO J. 20(11), 2923-2930. https://doi.org/10.1093/emboj/20.11.2923.

Thybert, D., Roller, M., Navarro, F.C.P., Fiddes, I., Streeter, I., Feig, C., Martin-Galvez, D., Kolmogorov, M., Janoušek, V., Akanni, W., Aken, B., Aldridge, S., Chakrapani, V., Chow, W., Clarke, L., Cummins, C., Doran, A., Dunn, M., Goodstadt, L., Howe, K., Howell, M., Josselin, A. A., Karn, R.C., Laukaitis, C.M., Jingtao, L., Martin, F., Muffato, M., Nachtweide, S., Quail, M.A., Sisu, C., Stanke, M., Stefflova, K., Van Oosterhout, C., Veyrunes, F., Ward, B., Yang, F., Yazdanifar, G., Zadissa, A., Adams, D.J., Brazma, A., Gerstein, M., Paten, B., Pham, S., Keane, T.M., Odom, D. T., Flicek, P., 2018. Repeat associated mechanisms of genome evolution and

function revealed by the *Mus caroli* and *Mus pahari* genomes. Genome Res. 28(4), 448-459. https://doi.org/10.1101/gr.234096.117.

Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., Vaillant, I., 2010. Stress-induced activation of heterochromatic transcription. PLoS Genet. 6(10), e1001175. https://doi.org/10.1371/journal.pgen.1001175.

Travers, A.A., Muskhelishvili, G., Thompson, J.M.T., 2012. DNA Information: from digital code to analogue structure. Philos. Trans. R. Soc. Lond. A. 370(1699), 2960–2986. https://doi.org/10.1098/rsta.2011.0231.

Trizzino, M., Kapusta, A., Brown C.D., 2018. Transposable elements generate regulatory novelty a tissue-specific fashion. BMC Genomics 19, 468. https://doi.org/10.1186/s12864-018-4850-3.

Tsay, Y.F., Frank, M.J., Page, T., Dean, C., Crawford, N.M., 1993. Identification of a mobile endogenous transposon in *Arabidopsis thaliana*. Science 260(5106), 342-344. https://doi.org/10.1126/science.8385803.

Tsuchiya, T., Eulgem, T., 2013. An alternative polyadenylation mechanism coopted to the *Arabidopsis RPP7* gene through intronic retrotransposon domestication. Proc. Natl. Acad. Sci. USA 110(37), E3535-E3543. https://doi.org/10.1073/pnas.1312545110.

Umeda, M., Ohtsubo, H., Ohtsubo, E. 1991. Diversification of the rice *Waxy* gene by insertion of mobile DNA elements into introns. Jpn. J. Genet. 66(5), 569-586. https://doi.org/doi: 10.1266/jjg.66.569.

Underwood, C.J., Choi, K., 2019. Heterogenous transposable elements as silencer, ehancers and targets of meiotic recombination. Chromosoma 128(3), 279-296. https://doi.10.1007/s00412-019-00718-4.

Usai, G., Mascagni, F., Giordani, T., Vangelisti, A., Bosi, E., Zuccolo, A., Ceccarelli, M., King, R., Hassani-Pak, K., Zambrano, L. C., Cavallini, A., Natali, L., 2020a. Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. Plant J. 102(3), 600-614. https://doi.org/10.1111/tpj.14635.

Usai, G., Mascagni, F., Vangelisti, A., Giordani, T., Ceccarelli, M., Cavallini, A., Natali, L., 2020b. Interspecific hybridisation and LTR-retrotransposon mobilisation-related structural variation in plants: A case study. Genomics 122(2), 1611-1621 https://doi.org/10.1016/j.ygeno.2019.09.010.

Usai, G., Vangelisti, A., Simoni, S., Giordani, T., Natali, L., Cavallini, A., Mascagni, F., 2021. DNA modification patterns within the transposable elements of the fig (*Ficus carica* L.) genome. Plants 10(3), 451. https://doi.org/10.3390/plants10030451.

van de Lagemaat, L.N., Gagnier, L., Medstrand, P., Mager, D.L., 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res. 15 (9), 1243-1249. https://doi.org/10.1101/gr.3910705.

Vangelisti, A., Mascagni, F., Giordani, T., Sbrana, C., Turrini, A., Cavallini, A., Giovannetti, M., Natali, L., 2019. Arbuscular mycorrhizal fungi induce the expression of specific retrotransposons in roots of sunflower (*Helianthus annuus* L.). PLoS ONE 14(2): e0212371. https://doi.org/10.1371/journal.pone.0212371.

Van Regenmortel, M.H.V., 2002. Reductionism and the search for structure-function relationships in antibody molecules. J. Mol. Recognit. 15(2), 240-247. https://doi.org/10.1002/jmr.584.

Van Regenmortel, M.H.V., 2004. Biological complexity emerges from the ashes of genetic reductionism. J. Mol. Recognit. 17(3), 145-148. https://doi.org/10.1002/jmr.674.

Van't Hof, A.E., Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby, A.C., Saccheri, I.J., 2016. The industrial melanism mutation in British peppered moths is a transposable element. Nature 534(7605), 102-105. https://doi.org/10.1038/nature17951.

Venner, S., Feschotte, C., Biémont, C., 2009. Transposable element dynamics: toward a community ecology of the genome. Trends Genet. 25(7), 317-323. https://doi.org/10.1016/j.tig.2009.05.003.

Ventimiglia, M., Pugliesi, C., Vangelisti, A., Usai, G., Giordani, T., Natali, L., Cavallini, A., Mascagni, F., 2020. On the trail of *Tetu1*: genome-wide discovery of *CACTA* transposable elements in sunflower genome. Int. J. Mol. Sci. 21(6), 2021. https://doi.org/10.3390/ijms21062021.

Vicient, C.M., Casacuberta, J.M., 2017. Impact of transposable elements on polyploid plant genomes. Ann. Bot. 120(2), 195-207. https://doi.org/10.1093/aob/mcx078.

Villasante, A., Abad, J.P., Planello, R., Mendez-Lago, M., Celniker, S.E., de Pablos B., 2007. *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. Genome Res. 17(12), 1909-1918. https://doi.org/10.1101/gr.6365107.

Volff, J.-N., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays 28(9), 913-922. https://doi.org/10.1002/bies.20452.

Vongs, A., Kakutani, T., Martienssen, R.A., Richards, E.J., 1993. *Arabidopsis thaliana* DNA methylation mutants. Science 260(5116), 1926-1928. https://doi.org/10.1126/science.8316832.

Wallau, G.L., Vieira, C., Loreto, E.S.L., 2018. Genetic exchange in eukariotes through horizontal transfer: connected by the mobilome. Mobile DNA, 9(1), 6, https://doi.10.1186/s13100-018-0112-9.

Walsh, A.M., Kortschak, R.D., Gardner, M.G., Bertozzi, T., Adelson, D.L., 2013. Widespread horizontal transfer of retrotransposons. Proc. Natl. Acad. Sci. USA 110(3), 1012-1016. https://doi.org/10.1073/pnas.1205856110.

Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglu, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kiburn, D., Rioux, J., Nusbaun, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M., Landr, E.S., 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280(5366), 1077-1082. https://doi.org/10.1126/science.280.5366.1077.

Warren, W.C., Hillier, L.W., Marshall Graves, J.A., Birney, E., Ponting, C.P., Grützner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T., Yang, S.P., Heger, A., Locke, D.P., Miethke, P., Waters, P.D., Veyrunes, F., Fulton, L., Fulton, B., Graves, T., Wallis, J., Puente, X.S., López-Otín, C., Ordóñez, G.R., Eichler, E.E., Chen, L., Cheng, Z., Deakin, J.E., Alsop, A., Thompson, K., Kirby, P., Papenfuss, A.T., Wakefield, M.J., Olender, T., Lancet, D., Huttley, G.A., Smit, A.F., Pask, A., Temple-Smith, P., Batzer, M.A., Walker, JA, Konkel, M.K., Harris, R.S., Whittington, C.M., Wong, E.S., Gemmell, N.J., Buschiazzo, E., Vargas Jentzsch, I.M., Merkel, A., Schmitz, J., Zemann, A., Churakov, G., Kriegs, J.O., Brosius, J., Murchison, E.P., Sachidanandam, R., Smith, C., Hannon, G.J., Tsend-Ayush, E., McMillan, D., Attenborough, R., Rens, W., Ferguson-Smith, M., Lefèvre, C.M., Sharp, J.A., Nicholas, K.R., Ray, D.A., Kube, M., Reinhardt, R., Pringle, T.H., Taylor, J., Jones, R.C., Nixon, B., Dacheux, J.L., Niwa, H., Sekita, Y., Huang, X., Stark, A., Kheradpour, P., Kellis, M., Flicek, P., Chen, Y., Webber, C., Hardison, R., Nelson, J., Hallsworth-Pepin, K., Delehaunty, K., Markovic, C., Minx, P., Feng, Y., Kremitzki, C., Mitreva, M., Glasscock, J., Wylie, T., Wohldmann, P., Thiru, P., Nhan, M.N., Pohl, C. S., Smith, S.M., Hou, S., Nefedov, M., de Jong, P.J., Renfree, M.B., Mardis, E.R.,

Wilson, R.K., 2008. Genome analysis of the platypus reveals unique signatures of evolution. Nature 453(7192), 175-183. https://doi.org/10.1038/nature06936.

Warren, I.A., Naville, M., Chalopin, D., Levin, P., Berger, C.S., Galiana, D., Volff, J.-N., 2015. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. Chromosom. Res. 23(3), 505-531. https://doi.org/10.1007/s10577-015-9493-5.

Waters, P.D., Dobigny, G., Waddell, P.J., Robinson, T.J., 2007. Evolutionary history of LINE-1 in the major clades of placental mammals. PLoS One 2, e158. https://doi.org/10.1371/journal.pone.0000158.

West, P., Li, Q., Ji, L., Eichten, S., Song, J., Vaughn, M.W., Schmitz, R.J., Springer, N.M., 2014. Genomic distribution of H3K9me2 and DNA methylation in a maize genome. PLoS One 9(8), e105267. https://doi.org/10.1371/journal.pone.0105267.

Wibowo, A., Becker, C., Marconi, G., Durr, J., Price, J., Hagmann, J., Papareddy, R., Putra, H., Kageyama, J., Becker, J., Weigel, D., Gutierrez-Marcos, J., 2016. Hyperosmotic stress memory in *Arabidopsis* is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. eLife 5, e13546. https://doi.org/10.7554/eLife.13546.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8(12), 973-982. https://doi.org/10.1038/nrg2165.

Wilden, A., 1980. System and Structure. Tavistock Publications Ltd, New York.

Williams, E.G., Auwerx, J., 2015. The convergence os systems and reductionist approaches in complex trait analysis. Cell 16(1), 23-32 https://doi.org/10.1016/j.cell.2015.06.024.

Wills, P.R., 2001. Autocatalysis, information and coding. Biosystems 60, 49-57. https://doi.org/10.1016/s0303-2647(01)00107-1.

Wills, P.R., 2014. Genetic information, physical interpreters and thermodynamics; the material-informatic basis of biosemiosis. Biosemiotics 7, 141-165. https://doi.org/10.1007/s12304-013-9196-2.

Wills, P.R., 2016. The generation of meaningful information in molecular systems. Phil. Trans. R. Soc. A 374, 20150066. https://doi.org/10.1098/rsta.2015.0066.

Wright, D.A., Voytas, D.F., 1998. Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana Ty3/gypsy* retrotransposons that encode envelope-like proteins. Genetics 149(2), 703-715.

Wu, S.C., Zhang, Yi., 2010. Active DNA demethylation: many roads lead to Rome. Nat. Rev. Mol. Cell Biol. 11(10), 750. https://doi.org/10.1038/nrm2950.

Wyatt, G.R., 1951. Recognition and estimation of 5-methylcytosine in nucleic acids. Biochem. J. 48(5), 581-584. https://doi.org/10.1042/bj0480581.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., van der Knaap, E., 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science 319(5869), 1527-1530. https://doi.org/10.1126/science.1153040.

Yu, C., Zhang, J., Peterson, T., 2011. Genome rearrangements in maize induced by alternative transposition of reversed *Ac/Ds* termini. Genetics 188(1), 59-67. https://doi.org/10.1534/genetics.111.126847.

Yu, J., Marshall, K., Yamaguchi, M., Haber, J.E., Weil, C.F., 2004. Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in *Saccharomyces cerevisiae*. Mol. Cell Biol. 24(3), 1351-1364. https://doi.org/10.1128/MCB.24.3.1351-1364.2004.

Zakrzewski, F., Schmidt, M., Van Lijsbetten, M., Schmidt, T., 2017. DNA methylation of retrotransposons, DNA transposons and genes in sugar beet (*Beta vulgaris* L.). Plant J. 90(6), 1156-1175. https://doi.org/10.1111/tpj.13526.

Zemach, A., McDaniell, E., Silva, P., Zilberman, D., 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328 (5980), 916-919. https://doi.org/10.1126/science.1186366.

Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., Zilberman, D., 2013. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153(1), 193-205. https://doi.org/10.1016/j.cell.2013.02.033.

Zhang, H., Lang, Z., Zhu, J.K., 2018. Dynamics and function of DNA methylation in plant. Nat. Rev. Mol. Cell Biol. 19(8), 489-506. https://doi.org/10.1038/s41580-018-0016-z.

Zhang, J., Yu, C., Krishnaswamy, L., Peterson, T., 2011. Transposable elements as catalysts for chromosome rearrangements. Methods Mol. Biol. 701, 315-326. https://doi.org/10.1007/978-1-61737-957-4_18.

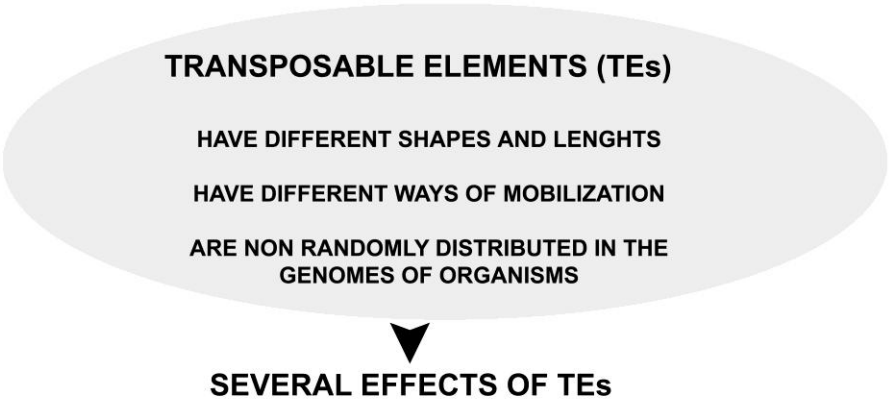
Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., Ecker, J.R., 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. Cell 126(6), 1189-1201. https://doi.org/10.1016/j.cell.2006.08.003.

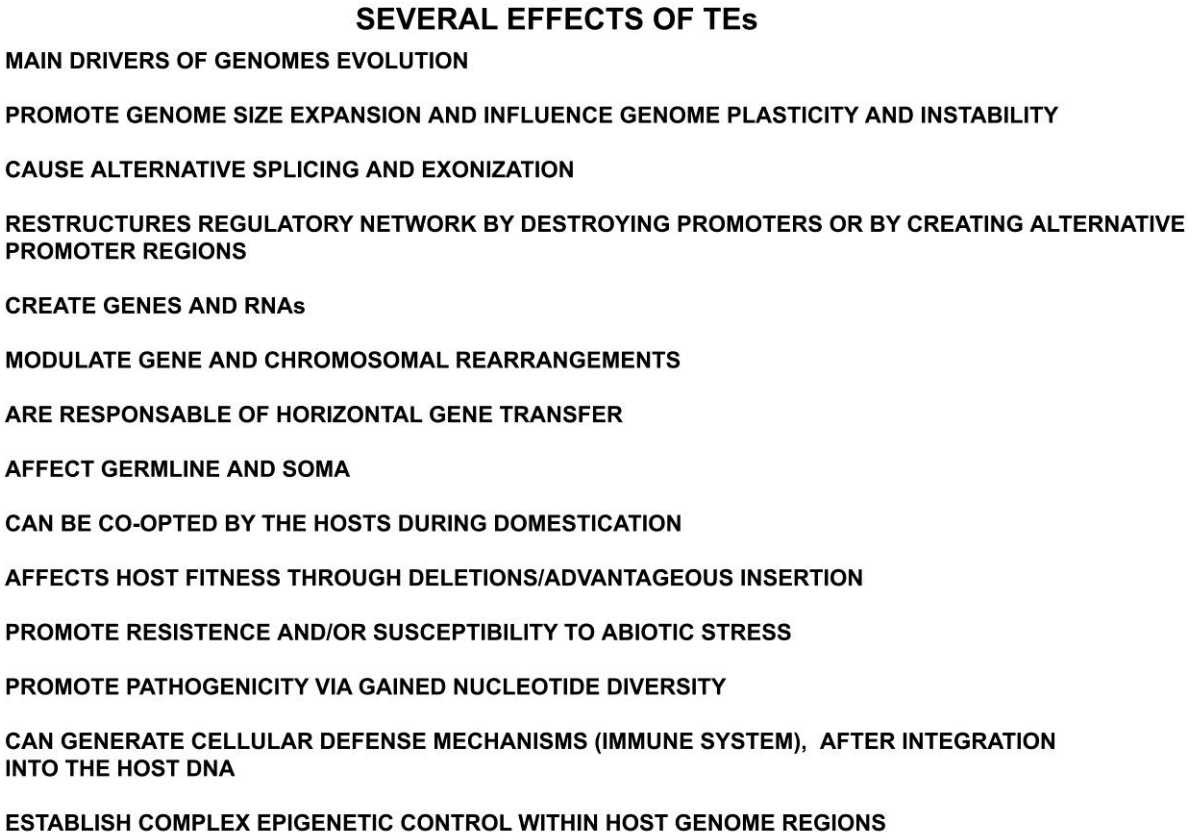
Zhao, H., Zhang, W., Chen, L., Wang, L., Marand, A.P., Wu, Y., Jiang, J., 2018. Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. Plant Physiol. 176(4), 2789-2803. https://doi.org/10.1104/pp.17.01467.

Zhu, J., Kapoor, A., Sridhar, V.V., Agius, F., Zhu, J.K. 2007. The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*. Curr. Biol.17(1), 54-59. https://doi.org/10.1016/j.cub.2006.10.059.

Zhu, J.K., 2009. Active DNA demethylation mediated by DNA glycosylases. Annu. Rev. Genet. 43, 143-166. https://doi.org/10.1146/annurev-genet-102108-134205.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., Henikoff, S., 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. 39(1), 61-69. https://doi.org/10.1038/ng1929.





1745 **Fig. 1.**

1746

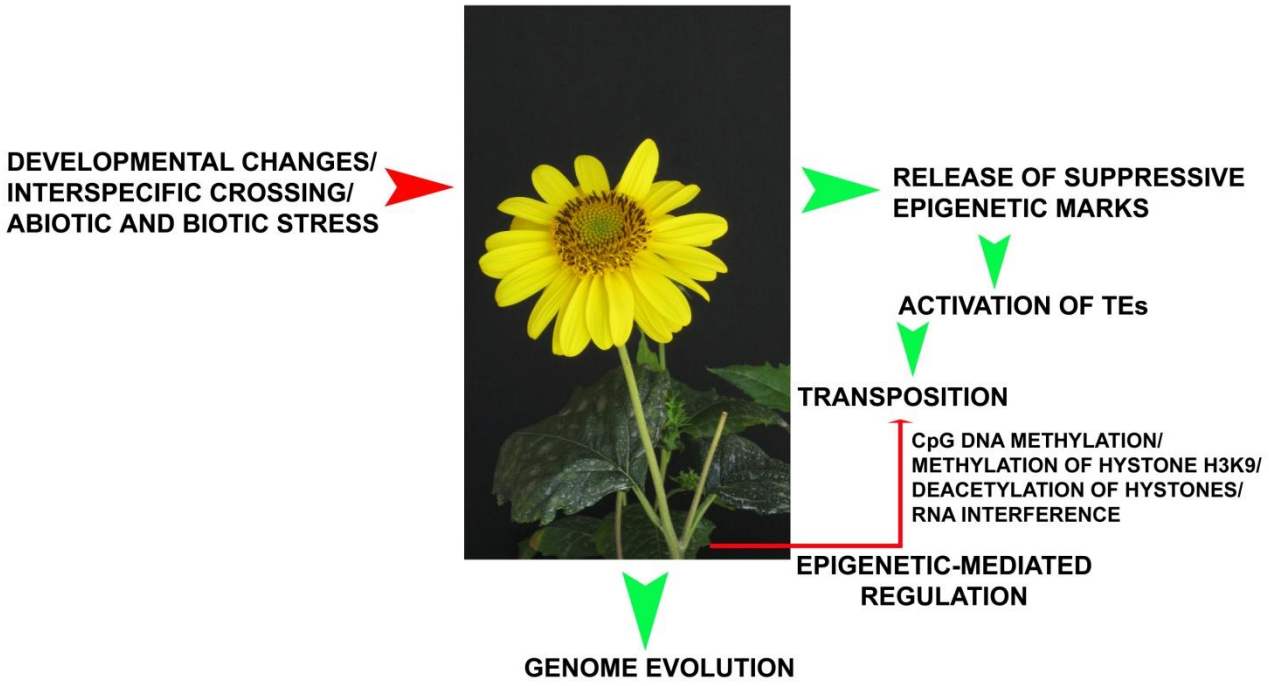

DEVELOPMENTAL CHANGES/
INTERSPECIFIC CROSSING/
ABIOTIC AND BIOTIC STRESS
RELEASE OF SUPPRESSIVE
EPIGENETIC MARKS
ACTIVATION OF TEs
TRANSPOSITION
CpG DNA METHYLATION/
METHYLATION OF HYSTONE H3K9/
DEACETYLATION OF HYSTONES/
RNA INTERFERENCE
EPIGENETIC-MEDIATED
REGULATION
GENOME EVOLUTION

Fig. 2.

1747