Gene 665 (2018) 155–166

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Review

Contribution of transposable elements in the plant's genome

Mahbod Sahebi[a,*], Mohamed M. Hanafi[a,b,c,**], Andre J. van Wijnen[d], David Rice[e], M.Y. Rafii[a], Parisa Azizi[f], Mohamad Osman[f], Sima Taheri[f], Mohd Faizal Abu Bakar[g], Mohd Noor Mat Isa[g], Yusuf Muhammad Noor[g]

[a] *Laboratory of Climate-Smart Food Crop Production, Institute of Tropical Agriculture and Food Security, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*
[b] *Laboratory of Plantation Science and Technology, Institute of Plantation Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*
[c] *Department of Land Management, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*
[d] *Mayo Clinic, Rochester, Minnesota, USA*
[e] *Department of Molecular Biology & Biotecnology, University of Sheffield, United Kingdom*
[f] *Department of Crop Science, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*
[g] *Malaysia Genome Institute, Jalan Bangi, 43000 Kajang, Malaysia*

ARTICLE INFO

Keywords:
Transposable elements
Gene expression
Plant development
TEs silencing
Stress conditions

ABSTRACT

Plants maintain extensive growth flexibility under different environmental conditions, allowing them to continuously and rapidly adapt to alterations in their environment. A large portion of many plant genomes consists of transposable elements (TEs) that create new genetic variations within plant species. Different types of mutations may be created by TEs in plants. Many TEs can avoid the host's defense mechanisms and survive alterations in transposition activity, internal sequence and target site. Thus, plant genomes are expected to utilize a variety of mechanisms to tolerate TEs that are near or within genes. TEs affect the expression of not only nearby genes but also unlinked inserted genes. TEs can create new promoters, leading to novel expression patterns or alternative coding regions to generate alternate transcripts in plant species. TEs can also provide novel *cis*-acting regulatory elements that act as enhancers or inserts within original enhancers that are required for transcription. Thus, the regulation of plant gene expression is strongly managed by the insertion of TEs into nearby genes. TEs can also lead to chromatin modifications and thereby affect gene expression in plants. TEs are able to generate new genes and modify existing gene structures by duplicating, mobilizing and recombining gene fragments. They can also facilitate cellular functions by sharing their transposase-coding regions. Hence, TE insertions can not only act as simple mutagens but can also alter the elementary functions of the plant genome. Here, we review recent discoveries concerning the contribution of TEs to gene expression in plant genomes and discuss the different mechanisms by which TEs can affect plant gene expression and reduce host defense mechanisms.

1. Introduction

Plants have tremendous growth flexibility, allowing them to continuously and rapidly adapt to environmental changes. Plants require genetic variation to tolerate both abiotic and biotic stress conditions. Consistent with the selfish DNA theory, TEs with extensive variation in their number, diversity and location in different chromosomes have been maintained in plants over evolutionary time (Feschotte and Pritham, 2007). Genetic variation ranging from single nucleotide polymorphisms to duplication of the entire genome and epigenetic variation contribute to the ability of plants to adapt to changing environments.

The genomes of most eukaryotes are mainly composed of non-coding sequence regions. These sequences include intergenic DNA, repetitive elements and introns. High copy numbers of transposable elements form a large portion of the repetitive elements in the eukaryotic genome. Transposable elements (TEs) are fundamentally different from genes (Saze and Kakutani, 2011) as recombination mostly occurs within genes and rarely at the level of TEs (Fu et al., 2002).

In contrast to genes that participate in a set of chromatin

Abbreviations: Transposable elements, (TEs); long terminal repeats, (LTRs); long/short interspersed elements, (LINEs/SINEs); Transportable long interspersed elements, (LINEs); RNA polymerase II, (Pol II); Long terminal repeat, (LTR); Repeat-induced point, (RIP); Transcription start sites, (TSSs)

[*] Corresponding author.

[**] Correspondence to: M. M. Hanafi, Laboratory of Climate-Smart Food Crop Production, Institute of Tropical Agriculture and Food Security, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia.

E-mail addresses: mahbod_sahebi@yahoo.com (M. Sahebi), mmhanafi@upm.edu.my (M.M. Hanafi).

https://doi.org/10.1016/j.gene.2018.04.050
Received 18 December 2017; Received in revised form 4 April 2018; Accepted 18 April 2018

modifications related to active transcription, such as histone H3 methylation at lysine 4 and histone acetylation, TEs are DNA sequences that are enriched for chromatin modifications associated with transcriptional silencing, such as DNA, histone (e.g., H3K27, H3K9) and CHG (where H is any nucleotide other than G) methylation (Saze and Kakutani, 2011). TEs have been found in the genome sequences of almost all eukaryotes and are capable of integrating elsewhere in a genome (Huang et al., 2012).

Transposons were discovered in maize (*Zea mays*) following the observation of unusual color patterns in leaves. (McClintock, 1948). Moreover, due to their large size, maize chromosomes can be easily visualized using a light microscope. Researchers discovered two loci, namely, dissociation (Ds) and activator (Ac) loci, that played roles in chromosome 9 breakage. Ds was found inside the breakage site, and Ac was found to be an essential locus for the activation of chromosome breakage. Ds and Ac were discovered to have variable locations among generations, identifying them as genetic elements able to undergo transposition. Transposable elements are known as selfish genetic factors because they are generally able to autonomously change locations within the plant genome through both copy-paste (retrotransposon) and cut-and-paste (DNA transposon) mechanisms.

2. LINEs and SINEs

There are three main classes of retro-transposons; long terminal repeats (LTRs), long interspersed elements (LINEs) and short interspersed elements (SINEs). Generally, retro-transposons use a copy-paste mechanism to amplify their DNA sequences, leading to the formation of large portions of repeats in eukaryotic genomes. Transportable LINEs are transcribed by RNA polymerase II and function as autonomous retro-transposon complexes that can be thousands of base pairs in length (Beck et al., 2011). The internal RNA polymerase II (Pol II) promoter of LINEs creates an mRNA-like capped and polyadenylated RNA transcript (Kroutter et al., 2009). Based on previous studies (Ostertag and Kazazian Jr, 2001; Dewannieux et al., 2003; Denli et al., 2015), an example of LINEs in the human genome is described herein to illustrate the copy-paste mechanism of retro-transposon elements.

SINE family members typically have a length of 85 to 500 base pairs (Kramerov and Vassetzky, 2011). Three different parts, the head (5′ end), body and tail (3′ end), compose a short interspersed element (SINE). The sequence of the head portion provides a substrate for the internal promoter (Pol III) and is used to classify SINEs into three different superfamilies based on their origin and consequent similarity to i) cellular Pol III genes encoding tRNAs, ii) 7SL RNA (such as human Alu elements), and iii) 5S rRNA (SINE3) (Kramerov and Vassetzky, 2011; Kapitonov and Jurka, 2003; Wicker et al., 2007). Together, LINEs and SINEs represent up to 30% of the sequences in the human genome and exhibit a nonrandom distribution (Lander et al., 2001). It has been reported that SINEs are typically located in gene-rich regions of the genome, while LINEs are mostly located in intergenic regions (Medstrand et al., 2002). LINEs and SINEs show different effects of the insertion of their large and small sequences, respectively, into genic regions: most SINEs regulate gene expression in eukaryotes, whereas LINEs can have negative effects following insertion into or near genes. The evolution of LINEs and SINEs has led to the remarkable diversity of eukaryote genomes through intra-chromosomal and inter-chromosomal recombination between allied LINEs or SINEs. This intra-chromosomal and inter-chromosomal recombination leads to genomic deletions and genomic rearrangements, respectively, creating homologous regions that function as hotspots in similar non-allelic recombination events and that provide robust storage of different sequences with varied functions, such as coding, regulatory and disruptive sequences (Sen et al., 2006; Feschotte, 2008). Both SINEs and LINEs support mRNA retro-transposition, which can contribute to genetic diversity among species and potentially within species in many ways. The resulting "retro-genes" situated near the functional region are not under selective pressure; therefore they have the potential to generate mutations or attain novel functions (Kaessmann et al., 2009).

Plant genome structure and function can be considerably affected by TE activities; for example, TEs can lead to gene disruption; genomic rearrangements that lead to the translocation, deletion and duplication of genetic elements; and changes to gene expression (Fedoroff, 2012; Lisch, 2013). The excessive activity of transposable elements is typically suppressed by genome defense mechanisms (Feng et al., 2010). DNA methylation, as a part of the plant genome defense mechanism, can suppress TE activity to avoid their negative effects on plant growth and development.

TE-induced mutations may be readily eliminated following their appearance in a population, with the probability of spreading depending partially on the effective population size. However, the genetic variability created by transposable elements may contribute to genome evolution in some cases. For instance, the insertion of a TE into FLOWERING LOCUS C (FLC) as the first intron of *Arabidopsis thaliana* results in an early flowering time, meeting the demand for vernalization in certain environments and affecting plant reproduction (Strange et al., 2011). It has been reported that transposable elements that are suppressed by plant genome-defense mechanisms may become reactivated during the exposure of the plant to environmental stress (Capy et al., 2000). Hence, the mobilization of TEs enables the formation of genetic variability that allows plant adaptability.

Plants are subject to both abiotic stresses, such as extreme weather conditions, and biotic stresses, such as pathogen stress. TEs play critical roles in launching genetic plasticity during plant-pathogen interactions (Lisch, 2013; Raffaele and Kamoun, 2012; Dong et al., 2015). A large number of transposable elements function in the response to stress by altering gene expression, and detailed studies of the distributions of genome-extensive TEs and their impact on genome evolution are needed. Although TEs do not prevent stress, they can alter gene expression in response to different environmental stresses. A high abundance of TE copies often hinders the generation of good-quality genome assemblies, with extensive fragmentation of TE-rich regions (Thomma et al., 2016). Chromatin interaction maps and long-read sequencing technologies have recently been used to reveal the genomic sequences of chromosomes and to establish uninterrupted genome assemblies, helping to identify the effects of TEs on genome evolution (Thomma et al., 2016; Gibriel et al., 2016).

The extensive diversity of TEs found among different genomes may facilitate genome descriptions, but TE diversity within genomes is poorly understood. Interactions are expected between coding and noncoding regions of a genome sequence that carry out necessary functions. Here, we consider the effects of TEs on gene expression and plant development under extreme growth conditions. In addition, we discuss potential mechanisms that allow TEs to affect gene expression and to act within coding sequences in roles other than insertional mutagenesis.

3. Transposable elements in the plant genome

Diverse types of TEs are found in plant genomes. Two main classes (I & II) are recognized based on transposition mechanism (Wicker et al., 2007). The first class of TEs comprises retro-transposons that use a copy-and-paste mechanism to transpose an RNA intermediate. These TEs include long terminal repeat (LTR) elements, short interspersed nuclear elements (SINEs) and long interspersed nuclear elements (LINEs). SINEs and LINEs are known as non-LTR retro-transposons. Retro-transposons are commonly found in the genomes of plants and mostly involve long terminal repeats (LTRs). They can be divided into five orders and subdivided into multiple superfamilies, such as Gypsy and Copia, according their structures in the plant genome (Wicker et al., 2007) (Kumar and Bennetzen, 1999). The second class of TEs includes DNA transposons, which contain terminal inverted repeats (TIRs) that use a cut-and-paste mechanism through a DNA intermediate for direct transposition. DNA transposons are also found in most plant