Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Duplication of host genes by transposable elements

Stefan Cerbin and Ning Jiang

The availability of large amounts of genomic and transcriptome sequences have allowed systematic surveys about the host gene sequences that have been duplicated by transposable elements. It is now clear that all super-families of transposons are capable of duplicating genes or gene fragments, and such incidents have been detected in a wide spectrum of organisms. Emerging evidence suggests that a considerable portion of them function as coding or non-coding sequences, driving innovations at molecular and phenotypic levels. Interestingly, the duplication events not only have to occur in the reproductive tissues to become heritable, but the duplicated copies are also preferentially expressed in those tissues. As a result, reproductive tissues may serve as the 'incubator' for genes generated by transposable elements.

Address
Department of Horticulture, 1066 Bogue Street, Michigan State University, East Lansing, MI 48824, USA

Corresponding author: Jiang, Ning (jiangn@msu.edu)

Current Opinion in Genetics & Development 2018, **49**:63–69

This review comes from a themed issue on **Genome architecture and expression**

Edited by **Kathleen Burns** and **Damon Lisch**

https://doi.org/10.1016/j.gde.2018.03.005

Introduction

The first examples of gene capture by transposable elements (TEs) were revealed more than 40 years ago, when genetic elements carrying drug resistance genes were found to be mobile in a genome or cause horizontal transfer among different bacteria [1,2]. In many cases, the resistance gene is carried by a composite element which consists of two independent mobile elements flanking the gene. For example, the Tn903 element contains two identical IS903 elements in reversed orientation at its termini, forming a 1 kb long terminal inverted repeat (TIR), surrounding the kanamycin resistance gene [3] (Figure 1). Although the Tn903 element moves as a single unit, each of the IS903 elements can move independently. In 1988, the first gene capture event in plants was reported, where the maize *Mu1/Mu2* elements were shown to carry a piece of sequence from the gene MRS-A [4]. *Mu1/Mu2* elements are members of the *Mutator*-like element (MULE) superfamily of DNA transposons, and most of them are associated with extended long TIRs, ranging from 100 to 800 bp [5]. Subsequently, MULEs carrying gene fragments were found in more organisms and called 'Pack-MULEs' [6–8]. Unless harboring additional nested insertions, Pack-MULEs are less than 5 kb in length. Since *Mu1/Mu2* elements contain long TIRs, it was proposed that they resemble the composite elements in bacteria, that is, each of the TIR is mobile independently. This gene capture would occur when a pair of TIR lands on both side of a gene, forming a Pack-MULE [4]. This hypothesis was tested recently, yet no transposition activity was detected with a single TIR [9••], suggesting MULEs and composite transposons in bacteria likely employ distinct mechanisms to acquire or duplicate host genomic sequences.

As sequencing technology advances, many further examples of gene duplication events by TEs, in a wide spectrum of organisms, were reported. Moreover, the availability of large amounts of transcriptome and proteome data facilitates the understanding of the expression and evolutionary trajectory of sequences duplicated by TEs. In this review, we discuss some of the latest progress in this area.

Different type of duplications and widespread presence of gene duplication by TEs

TEs are categorized into two classes based on their transposition mechanism. Class I, retrotransposons, utilize an RNA intermediate for transposition. These including long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons [10]. Class II, DNA transposons, transpose via a DNA intermediate [11]. Over 10 superfamilies of DNA transposons have been identified hitherto [12–14]. For most TEs, the transposition is accompanied by the duplication of a small piece of genomic sequence, which is called a target site duplication (TSD). Retrotransposons convert gene transcripts into cDNAs and then integrate the retrocopy, with or without being attached to a recognizable TE sequence, into the genome. Non-LTR elements generate retrocopies by reading-through their flanking sequences, which is called transduction [15,16] (Figure 1). In this case, the retrocopy is attached to the 3′ end of the element. An LTR element may carry the retrocopy or gene fragment between the two LTRs of a single element (Figure 1), and this is likely achieved via a template switch during cDNA synthesis of element sequence [17••]. The retrocopies are distinguished from their parental genes by the presence of poly-A tails, TSDs and most importantly, lack of introns.

Figure 1

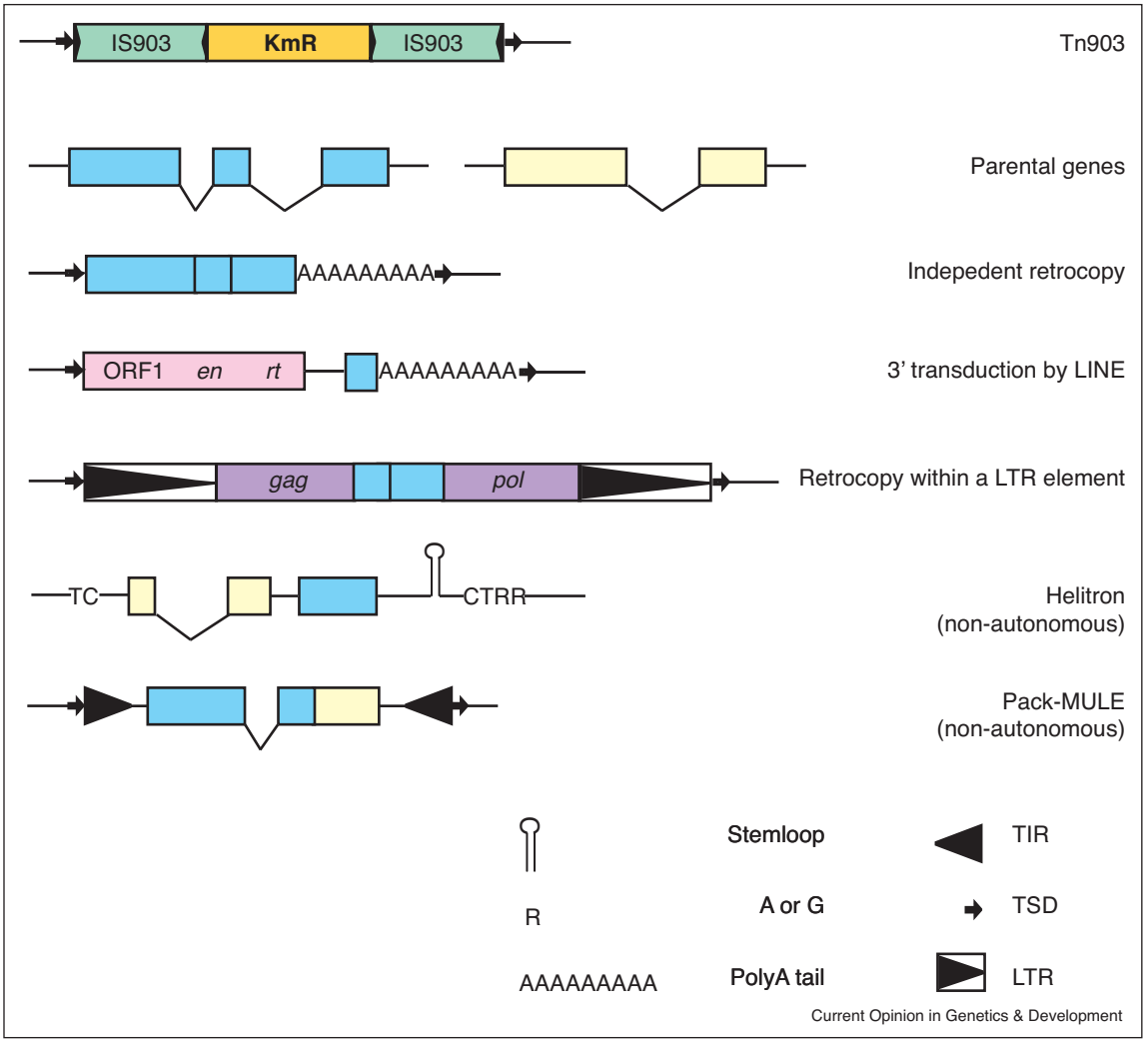

Diagram of different types of gene duplications generated by transposable elements. Colored boxes indicate exons or coding regions. Black triangles indicate terminal inverted repeat or long terminal repeat. Small black arrows indicate target site duplications. Introns are shown by 'V' shape line connecting exons. Other sequences are shown by black lines. KmR: kanamycin resistance gene; ORF1, en and rt, proteins encoded by long interspersed element (LINE, belong to non-LTR retrotransposons, en and rt are encoded by ORF2); gag and pol, proteins encoded by LTR retrotransposons. *Helitron* elements start with 'TC' at the 5′ termini and end with 'CTRR' (R = A or G) at the 3′ end. Moreover, there is a stem–loop close to the 3′ end of *Helitron*s.

If a retrocopy is functional, it is called a retrogene, otherwise it is called a retropseudogene. The mechanism of gene duplication by DNA transposons is unclear, and it remains unknown whether the duplication is associated with transposition. However, since introns and regulatory regions are retained, it is likely through a DNA intermediate [6]. As the two classes of transposons are associated with distinct transposition mechanisms, it is not surprising that they duplicate genes via different mechanisms.

In humans, L1, a non-LTR retrotransposon family, has duplicated 1% of the genome [18], resulting in a total of 18 700 retrocopies in human [19]. Based on RetrogeneDB [20], retrocopies have been identified from 62 animal and 37 plant species, covering the majority of evolutionary clades, with substantial amounts of LTR and non-LTR elements involved. By contrast, studies regarding gene duplication by DNA transposons are less comprehensive, and so far it has only been reported in few plant and animal genomes (see review in [5,21]). However, this does not mean DNA transposons duplicate genes less frequently; instead, emerging evidence suggests that DNA transposons more frequently carry genes or gene fragments than retrotransposons. For example, in rice, 87% of the members in the Os0037 MULE family harbor gene sequences in their internal regions [22]. In total, rice Pack-MULEs have duplicated sequences from 1500 parental genes. If Pack-MULE activity has been