

## PRIMER

# Transposable elements and microevolutionary changes in natural populations

GEORGI BONCHEV\*† and CHRISTIAN PARISOD\*

\*Laboratory of evolutionary botany, Institute of biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland, †Institute of plant physiology and genetics, Bulgarian academy of sciences, G. Bonchev Street, Bldg 24, Sofia 1113, Bulgaria

## Abstract

Transposable elements (TEs) usually represent the most abundant and dynamic fraction of genomes in almost all living organisms. The overall capacity of such 'junk DNA' to induce mutations and foster the reorganization of functional genomes suggests that TE may be of central evolutionary significance. However, to what extent TE dynamics drive and is driven by the evolutionary trajectory of host taxa remains poorly known. Further work addressing the fate of TE insertions in natural populations is necessary to shed light on their impact on microevolutionary processes. Here, we highlight methodological approaches (i.e. transposon displays and high-throughput sequencing), tracking TE insertions across large numbers of individuals and discuss their pitfalls and benefits for molecular ecology surveys.

**Keywords:** adaptation, genome dynamics, high-throughput sequencing, retrotransposons, speciation, transposon displays

Received 10 April 2013; revision received 31 May 2013; accepted 4 June 2013

## Introduction

Understanding the evolutionary ecology of taxa requires molecular mechanisms underlying variation and their ecological implications to be fully integrated. Our increasing knowledge of genome sequences already described a diversity of genome architectures and led to the concept of an evolutionarily dynamic genome (Lynch 2007; Koonin 2009). However, to what extent genome reorganization occurring at rapid rates influence and is influenced by the evolutionary trajectories of populations remains poorly known.

The discovery of transposable elements (TEs; i.e. DNA fragments from a few dozen of bp to 25 kb, having the ability to move within genomes; Box 1 and 2) entirely changed our appreciation of the stability of the genome (McClintock 1984). The functional and evolutionary impacts of TEs remain controversial today (e.g. Doolittle 2013). Although TEs can generally be appraised as parasites filling genomes with 'junk DNA' (Doolittle & Sapienza 1980; Orgel & Crick 1980), their biology has been primarily assessed by the phenotypic changes they induce through chromosomal rearrangements and interactions with coding sequences (Box 3). Selfish TEs indeed foster considerable variation that may influence

the evolution of their host taxa (Kidwell & Lisch 2001; Biémont & Vieira 2006).

With the advent of large-scale DNA sequencing, the last decade has offered a deeper understanding of the diversity and abundance of TEs (Box 1). It is now clear that eukaryote genomes comprise much more than sets of genes, being populated with large fractions of TEs (e.g. about 65% of the human genome, de Koning *et al.* 2011; and more than 80% of the large genomes of cereals, Li *et al.* 2004). Genome biologists are accordingly inclined to emphasize on the structural and functional significance of TEs, whereas a prevalent view among evolutionary biologists seems to be that TE insertions – in their vast majority – are nearly neutral and unlikely to have a strong evolutionary impact. This may be true as host genomes evolved mechanisms repressing TE activity and given that selection in large populations may efficiently purge deleterious insertions, despite high mutational pressure (Box 2). All TEs certainly do not have a significant role, but we are still largely ignorant about their quantitative impact on evolution. Furthermore, the common practice of filtering out TE sequences from analyses makes their evolutionary consequences less probably to be assessed. Although challenging, we argue that taking TEs into consideration when surveying natural populations may shed further light on their evolutionary impact. We thus suggest directions to implement such work in the agenda of molecular ecologists.

Correspondence: Christian Parisod, Fax: +41 (0)32 718 3001; E-mail: christian.parisod@unine.ch

**Box 1 A diverse community of transposable elements (TEs) in eukaryotes**

TEs are mobile sequences having their own life cycle within host genomes (Fig. 1). Two main types of TEs fundamentally differ in their mechanism of transposition (Wicker *et al.* 2007): Class I elements (or retrotransposons) move via 'copy and paste' mechanisms using RNA intermediates, whereas Class II elements (or DNA transposons) move via 'cut and paste' mechanisms through DNA intermediates. An important lower level of the hierarchical TE classification is the family. TE families are composed of individual insertions (i.e. particular TE copies at specific chromosomal loci) sharing sequence similarity (i.e. at least 80% identity over 80% of the internal regions and/or the terminal repeats) and regulation. Insertions of a given TE family thus forms a population of related sequences evolving within a given genome (thus within a population of host genomes).

**Types of TEs**

- 1 Class I retrotransposons transcribe RNA intermediates from genomic copies that are reverse-transcribed into a double-stranded DNA and integrated into a new position. Several daughter copies can be produced from a mother copy and insert throughout the genome, forming a major fraction of large genomes.
  - 1.1 Long terminal repeats (LTR) retrotransposons range from a few hundred base pairs up to 25 kb and are similar in structure to retroviruses (except that functional envelope genes are lacking). They typically encode the enzymes necessary for their transposition (incl. reverse transcriptase and integrase) and are delimited by LTR sequences harbouring promoter regions together with regulatory motifs ruling their transcription. Some LTR retrotransposons (sometimes described as terminal repeat in miniature, TRIM, or large retrotransposons derivatives, LARD) present nonfunctional ORFs and thus are nonautonomous, relying on autonomous elements for their transposition. LTR retrotransposons have been reported in all eukaryote genomes and are especially predominant in plants.
  - 1.2 Long interspersed nuclear elements (LINEs) can be up to several kb, presenting RNA polymerase II promoters at their 5' end, coding ORFs (incl. reverse transcriptase) and polyA tails at their 3' end. LINEs are found in all eukaryotes, but seem less common among unicellular ones. They typically predominate over LTR retrotransposons in animals.
  - 1.3 Short interspersed nuclear elements (SINEs) range from 80 to 500 bp. They derive from polymerase III transcripts and present corresponding internal promoters supporting their expression. SINEs, lacking reverse transcriptase, are nonautonomous and rely on LINEs for transposition. SINEs are found in all eukaryotes.
- 2 Class II elements typically excise from the donor site and conservatively transpose to another chromosomal location. The transposition of several DNA transposons is however coupled with replication. These TEs create duplication of 2 to 11 bp at the target site, and their excision thus leaves footprints in the genomes. DNA transposons are present in low–moderate-copy numbers in almost all eukaryotes.
  - 2.1 Terminal inverted repeats (TIRs) transposons range from few dozen to several hundreds of base pairs. They typically encode a transposase recognizing and cutting both DNA strands at the end of TIRs. Some TIR transposons (often referred as miniature inverted elements, MITEs) lack coding sequences and thus are nonautonomous, relying on autonomous elements for their transposition. DNA transposons have been reported in all eukaryote genomes, but also in prokaryotes in simple forms called insertion sequences.
  - 2.2 Recently described TEs such as Helitrons and Mavericks transpose through a replicative process involving only one strand. Not much is yet known about these large elements (up to 20 kb), but they often present several ORFs, including gene fragments captured from the host genome. Helitrons have been reported in plants, metazoa and fungi and replicate via a rolling circle mechanism. Mavericks seem to undergo extrachromosomal replication without RNA intermediates and have been described in diverse eukaryotes, except plants.

**TEs as evolutionary capacitors in natural populations**

The overall capacity of TEs to produce a wide diversity of mutations of significant impact advocates for their prominent role in fostering adaptive changes to their host genomes (Box 3). Despite the accumulation of case studies reporting evidence of circumstantial benefits

from specific TE insertions in natural populations (e.g. Shimizu *et al.* 2008; Wood *et al.* 2008; Paris & Despres 2012), limited comprehensive work on the adaptive potential of TEs has been conducted so far. Based on a suit of studies addressing adaptive TEs, González *et al.* (2010) showed that at least 32% of the putatively adaptive insertions screened in natural population of

## Box 2 Transposition rates and TE regulation

The vast majority of TE insertions within genomes are either defective (e.g. truncated) or epigenetically silenced, and only a minority of functional copies can be transcribed and effectively transpose (i.e. are active). Little is known about the absolute rate of transposition, but estimates from mutation accumulation experiments in *Drosophila melanogaster* reported from  $10^{-5}$  to  $10^{-3}$  insertions/copy/generation (e.g. Maside *et al.* 2000, 2001). This rate seems to reach  $10^{-2}$  after heat shock and up to  $10^{-1}$  after dysgenic crosses (Seleme *et al.* 1999). Insertion rates of a LTR retrotransposon about  $10^{-6}$  to  $10^{-4}$  have been reported in yeast (Garfinkel 2005). In plants, a DNA transposon has been estimated to transpose at a rate as high as 1 (Alleman & Freeling 1986).

Our understanding of the factors controlling the expression and transposition of TEs is growing and suggests that quiescent TEs increase their transposition rate under stress conditions or following hybridization (Grandbastien 1998; Capy *et al.* 2000). For instance, Tnt1 retrotransposons from tobacco present regulatory regions with specific motifs that are commonly observed in genes induced by drought, anaerobic conditions or oxidative stresses (Grandbastien *et al.* 2005). Similar examples from wheat retrotransposons under light and salinity stresses have been recently reported (Woodrow *et al.* 2010). Enhanced transcription of such TEs in response to specific environmental signals could thus boost transposition.

Transcriptional and post-transcriptional silencing mostly regulates the activity of TEs in plants, fungi and animals (reviewed in Castel & Martienssen 2013). RNA interference pathways involving TE-derived small interfering RNAs and PIWI-interacting RNAs indeed target homologous TE sequences and interact with them to repress their transcription. Physiological stresses may result in epigenetic repatterning inducing the activation of TEs. Gametogenesis in both plants and animals induces the repatterning of epigenetic marks, and gene flow between genomes enclosing divergent TEs may lead to genome instability by activating specific TEs (i.e. hybrid dysgenesis; Parisod & Senerchia 2013).

*Drosophila melanogaster* presented a distribution consistent with selection by contrasted temperature and rainfall regimes. Most of these TE insertions, belonging to multiple TE families, were located nearby functional genes that may account for the surmised phenotypic changes. Insights offered by such reverse population genomic approaches suggest that TEs represent an important source of adaptive variation, but such surveys should be undertaken in various species to reliably estimate the impact of TEs on the evolutionary ecology of species.

In addition to transposition, ectopic recombination among inserted TEs supports genome changes through chromosomal rearrangements and sequence deletion (Box 3, §3.4). The balance between transposition and deletion rules the ups and downs of genome sizes (Vitte & Panaud 2005). This is nicely illustrated by a wild relative of rice (*Oryza australiensis*) having doubled its genome size in ca. three million years under the influence of TEs (Piegu *et al.* 2006), whereas the genome of *Arabidopsis thaliana* predominantly presents truncated TEs and shrunk by ca. 30% over the last 10 million years (Hu *et al.* 2011). TE content largely explains patterns of the C-value paradox, but further work addressing the effects of population demography and host mating system on the dynamics of TEs in closely related taxa is necessary to disentangle the factors driving the evolutionary trajectories of genomes (Whitney *et al.* 2010).

TEs also drives high sequence turnover across functional genomes. In particular, genomes regularly show blocks of nested TE copies that can be greatly polymorphic among closely related taxa (e.g. Wang & Dooner 2006). For instance, the centromeres of *Arabidopsis thaliana* are significantly differentiated from *A. lyrata* in relation to the proliferation of a LTR retrotransposon in the latter species (Tsukahara *et al.* 2012). Similarly, the different lines of maize revealed highly contrasted arrangements of TE insertions across their intergenic space, and these structurally divergent genome blocks were shown to be recombinationally inert (He & Dooner 2009). TE dynamics thus determines fast-evolving regions, showing reduced recombination (Melamed-Bessudo & Levy 2012). Mapping of loci associated with seed and pollen sterility of crosses within the circumpolar *Draba nivalis* suggested that underdominant insertions of LTR retrotransposons might underlie cryptic speciation (Grundt *et al.* 2006; Gustaffsson & Parisod unpublished results). In wild barley (*Hordeum spontaneum*), individual growing in dry environments showed a higher number of full-length copies and fewer truncated insertions of a LTR retrotransposon than those growing a few dozen metre apart in less stressful habitats (Kalendar *et al.* 2000). The high turnover of TE fractions thus generated divergent genome arrangements that were closely matching the ecogeographical distribution of gene pools, suggesting that TEs sustained incipient speciation despite gene flow.

**Box 3 TE-induced mutations affecting ecologically relevant phenotypes**

TEs can affect genes and phenotypes through multiple mechanisms, depending on the TE itself and its insertion site. Altering coding sequences or their regulatory elements, TEs can directly affect phenotypes, but TEs also promote epigenetic changes impacting nearby genes and their products. Accordingly, TEs foster natural genetic engineering and have been compared with the evolutionary toolkit of the genome (Schmidt & Anderson 2006; Shapiro 2010). The few illustrative cases discussed below are all but exhaustive and readers will find additional examples of TE-induced mutations in recent reviews (e.g. Sela *et al.* 2010; Hua-Van *et al.* 2011; Fedoroff 2012; Lisch 2013).

**3.1 TE-induced loss-of-function**

The most obvious TE-induced change is gene disruption leading to observable loss-of-function (Fig. 2a). For instance, insertion of a TIR transposon into the *rugosa* locus encoding a starch-branching enzyme in pea resulted in the wrinkled seed studied by Mendel (Bhattacharyya *et al.* 1990). To date, approximately 65 diseases directly caused by TE insertions have been documented in humans (O'Donnell & Burns 2010).

**3.2 TE-induced regulatory changes**

TEs can directly influence gene expression through disruption of regulatory motifs (repressors or enhancers) or insertion of new regulatory information (Fig. 2b). TE-induced restructuring of the architecture of cis-promoter regions of genes has been regularly reported (Feschotte 2008). It is well illustrated by insertions of DNA transposons disrupting the promoter and reducing the transcription of the heat-shock protein 70 in natural lines of *Drosophila melanogaster*, altering thermotolerance as well as female reproductive success (Lerman *et al.* 2003; Lerman & Feder 2005). TEs can also bring their own regulatory sequences next to coding sequences and thereby influence gene expression (e.g. Jordan *et al.* 2003). For instance, insertion of a LTR retrotransposon upstream of the *Ruby* gene in blood orange has been shown to provide a new promoter controlling gene expression in response to cold and resulting in the red coloration of the fruit flesh (Butelli *et al.* 2012). Noticeably, deletion of the TE insertion in some derived varieties further enhanced the promoter action of the remaining LTR. Similarly, insertions of a LTR retrotransposon into the promoter of the cytochrome P450 gene *Cyp6 g1* of *Drosophila melanogaster* brought regulatory sequences rewiring and increasing the gene expression in tissues detoxifying a variety of insecticides, thereby conferring adaptive resistance (McCart & Ffrench-Constant 2008).

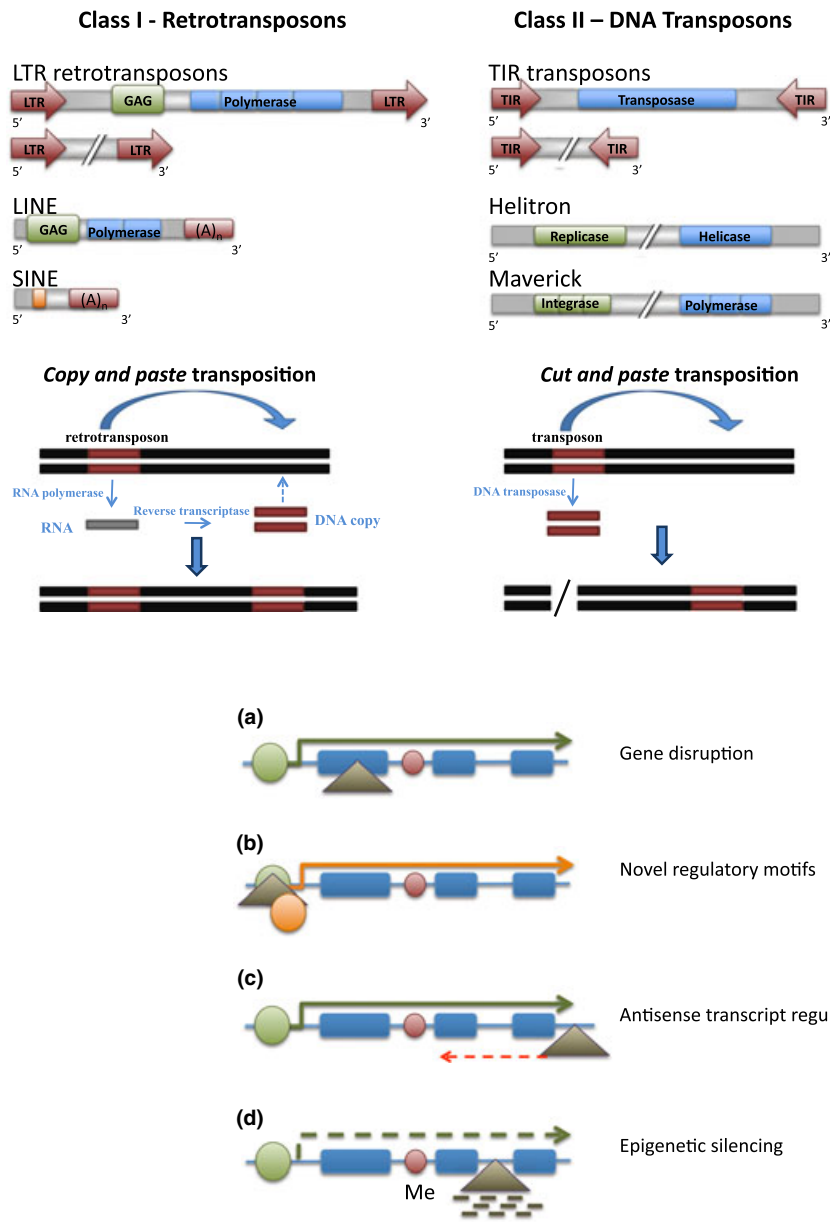
It is worth noting that gene expression is not only controlled by proximal promoters, but can also be influenced by TEs inserting at the 3' end of coding sequences or at relatively distant sites. For instance, variation in flowering time in maize is tightly associated with a MITE insertion disrupting a conserved noncoding region (*Vgt1*) located 70 kb away from the *AP2* transcription factor actually regulating this trait (Salvi *et al.* 2007). Similarly, overexpression of *tb1*, a gene repressing branching in cultivated maize, is controlled by the enhancer activity of an LTR retrotransposon insertion located 60 kb away from the gene (Studer *et al.* 2011).

**3.3 TE-induced epigenetic effects**

Genes located in the vicinity of TE insertions are probably methylated following the action of small RNAs targeting and repressing the expression of TEs. Although the structure of the protein-coding region remains unaltered, nearby TEs thereby produce stable epialleles that may be of evolutionary relevance (Slotkin & Martienssen 2007; Wang *et al.* 2013). The probability for a gene to be methylated and to show reduced expression seems to depend on the distance to the nearest TE. Such effects over more than 3000 bp have been regularly reported in *Arabidopsis* species (Hollister *et al.* 2011), indicating that TEs may indirectly influence gene expression over relatively large distances. For instance, insertion of a LTR retrotransposon near the *agouti* gene in mice produces an outward-reading antisense transcript that extends into the gene and interferes with the sense transcript (Morgan *et al.* 1999). The resulting variation in chromatin state and DNA methylation at this locus is determined by the epigenetic status of the retrotransposon that ultimately influences the gene transcript level and the colour of the mice coat (Fig. 2c). Similarly, early flowering of the Ler ecotype of *Arabidopsis thaliana* is controlled by a DNA transposon insertion in the first intron of *FLC* (a gene delaying flowering). This insertion is indeed targeted by TE-derived siRNAs resulting in the silencing of the gene, whereas ecotypes lacking this insertion show normal *FLC* expression and late flowering (Liu *et al.* 2004; Fig. 2d).

**3.4 TE-induced genome restructuring**

The repeated nature of TEs makes them efficient substrate for ectopic recombination. TEs are thus known to be responsible for deletions, insertions, inversions, translocations and duplications and to generally promote chromosomal instability (Gray 2000). Most types of genetic changes discussed above may thus occur through positional effects following chromosomal rearrangement supported by TEs.



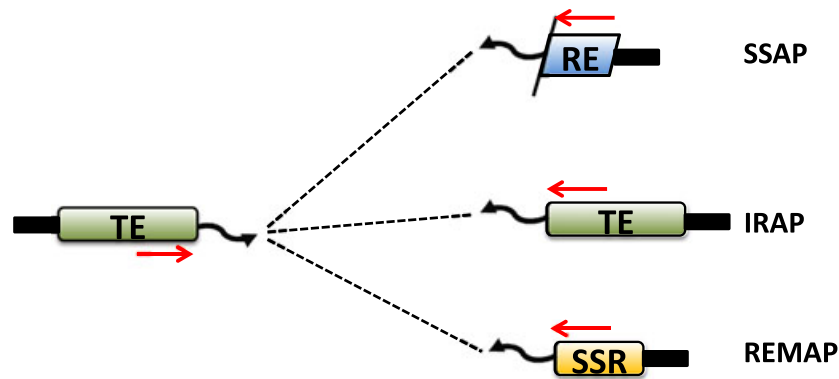
**Fig. 2** Effects of transposable element (TE) insertions on gene structure and function. (a) Insertion of TE into the coding regions of a gene can disrupt or abolish its function as illustrated by wrinkled-seed character of pea described by Mendel. (b) TEs can also provide novel promoter functions when inserted into regulatory sequences of a gene, altering its normal expression level. TE-derived expression of a gene responsible for the flesh colour of blood oranges is illustrated. (c) Transcription levels of genes can be interfered by antisense transcription from adjacent TE insertions as observed for the *agouti* colour gene in mice. The strength of the interference depends on the epigenetic state and expression level of the TE. (d) siRNA-controlled methylation of TE insertions can influence the expression of nearby genes as illustrated by the silencing of the *FLC* locus associated with the repression of an intronic TE insertion, speeding flowering time in *Arabidopsis thaliana*. Gene exons are depicted as blue rectangles and TE insertions as red triangles; promoters of genes and TEs are shown as green and orange circles, respectively. Arrowheads indicate gene transcription.

Convincing evidence of functional and/or adaptive impacts of TE insertions have been reported (Box 3), but may only represent the tip of the iceberg. As a whole, TEs emerge not only as chief producers of evolutionary significant genetic variation, but also as pivotal elements

further sheltering it from homogenizing gene flow by reducing recombination (Abbott *et al.* 2013). However, to what extent TEs account for microevolutionary processes such as rapid adaptive evolution in natural populations or speciation remains essentially to be addressed. Empir-

**Fig. 1** Structure of the different types of transposable elements and their schematic mechanisms of transposition. The ORF of long terminal repeat (LTR) retrotransposons is indicated here as GAG and Polymerase, although it typically codes for the several enzymes necessary for proper transposition such as the integrase, the RNase H and the reverse transcriptase domains. The orange box in SINEs represents the diagnostic tRNA-related region. TIR = terminal inverted repeats.





**Fig. 3** Main transposon displays tracking transposable element (TE) insertions and surveying insertional polymorphisms. The techniques rely on simultaneous PCR amplification of sequences between copies of a candidate TE and adjacent genomic regions. The targeted genomic regions differ among methods: other TE copies for inter-retrotransposon amplified polymorphism (IRAP), microsatellite loci (SSR) for retrotransposon–microsatellite amplified polymorphism (REMAP) or restriction sites (RE) for sequence-specific amplified polymorphism (SSAP). Red arrowheads indicate the primers.

ical data exploring the diversity and distribution of TE insertions through comparative genomics and population genetics are expected to provide a general view of (adaptive) evolution. Technological advances now make it possible to assess the role of TEs as capacitors of microevolutionary changes by shifting the focus from individual host genomes (i.e. the main concern of genome biologists) to populations (i.e. the main concern of evolutionary biologists).

### Selecting candidate TEs for molecular ecology surveys

Surveys of natural variation of TE insertions can hardly embrace the full diversity of TEs at once, and focusing on candidate TE families is necessary. Selecting candidate TEs for surveys of natural populations probably represents a critical step that depends on the specific objectives of the study. It has to rely on both the biology and the potential distribution of the TEs. In particular, tracking insertions of active TE families that potentially have generated abundant variation over ecological times seems a promising approach to assess the impact of TE dynamics on microevolutionary changes. Furthermore, TEs also differ by their size and distribution across genomes (Dolgin & Charlesworth 2008), and one may optimize the chance of obtaining insightful results by focusing on insertions of appropriate TE families. TEs are generally more abundant in the heterochromatin, centromeres and/or telomeres. In particular, high-copy-number families tend to form clustered (i.e. adjacent insertions) and nested (i.e. insertions within one another) arrangements in intergenic regions of large genomes (e.g. Baucom *et al.* 2009). In contrast, insertions of short

TEs such as MITEs, LINEs and SINEs, but also low-copy-number LTR retrotransposons, are observed all along the chromosomes and seem overrepresented near or within genes in plants and in humans (i.e. mostly in introns, in 5' or 3' UTR as well as flanking regions; Wright *et al.* 2003; Majewski & Ott 2002).

Practically, the most commonly used criterion to select TEs under study probably is to rely on information from model species and available databases (Table 1). Closely related species indeed have similar TE content and knowledge from model species can thus be transferred to wild relatives to a certain extent. As an example, the active LTR retrotransposon BARE-1, isolated from barley, has been shown to have close active homologues in other cereal genomes such as wheat, rye and oat, allowing to track polymorphic insertions in these species (e.g. Vicient *et al.* 2001; Saeidi *et al.* 2008). However, this is all but a universal feature as active TEs usually diverge relatively quickly. Accordingly, conserved regions of TEs used to design markers (see below) may either lack among related taxa or target ancient insertions of quiescent TE lineages.

Given the dearth of available information for species of ecological interest, characterizing TEs from the focal nonmodel species may represent a worthwhile investment. Isolation of novel TEs from taxa without genomic information can be achieved through the amplification and sequencing of a representative pool of heterogeneous TEs within a given species using degenerate primers matching conserved domains (e.g. the reverse transcriptase genes). Nonconserved regions of TEs can then be characterized through a genome-walking approach using TE specific and adapter primers (Kalendar *et al.* 2011). Such procedures have been successfully

**Table 1** Main databases devoted to transposable elements (TEs)

Abbreviation	Website	Comment
RepBase*	<a href="http://www.girinst.org/">http://www.girinst.org/</a>	Curated, nonredundant TE database from different eukaryotic species
GyDB 2.0†	<a href="http://gydb.org">http://gydb.org</a>	Database of Gypsy, Retroviridae, Copia, Bel/Pao LTR retroelements and the Caulimoviridae pararetroviruses of plants
SINEBase‡	<a href="http://sines.eimb.ru/">http://sines.eimb.ru/</a>	Database of SINEs of eukaryotes
TranspoGene	<a href="http://transpogene.tau.ac.il/">http://transpogene.tau.ac.il/</a>	TE within protein-coding genes of human, mouse, chicken, zebrafish, fruit fly, nematode and sea squirt
PlantRepeats	<a href="http://plantrepeats.plantbiology.msu.edu">http://plantrepeats.plantbiology.msu.edu</a>	Collections of TE insertions from <i>Hordeum</i> , <i>Oryza</i> , <i>Sorghum</i> , <i>Triticum</i> , <i>Glycine</i> , <i>Lotus</i> , <i>Medicago</i> , <i>Zea</i> , Brassicaceae, Solanaceae
TREP	<a href="http://wheat.pw.usda.gov/ITMI/Repeats">http://wheat.pw.usda.gov/ITMI/Repeats</a>	Curated collection of repetitive DNA sequences from Triticeae species (wheat, barley, rye, oat and their wild relatives)
Maizetdb	<a href="http://maizetdb.org/~maize/">http://maizetdb.org/~maize/</a>	Database of TEs from maize
Retroryza	<a href="http://retroryza.fr">http://retroryza.fr</a>	Database of LTR Retrotransposon from rice
SoyTEdb	<a href="http://soybase.org/soytedb/">http://soybase.org/soytedb/</a>	Database of TEs from soya bean and other legumes
RepPop	<a href="http://csbl.bmb.uga.edu/~ffzhou/RepPop/">http://csbl.bmb.uga.edu/~ffzhou/RepPop/</a>	Database of TEs from popular
DMEL	<a href="http://www.fruitfly.org/p_disrupt/TE.html">http://www.fruitfly.org/p_disrupt/TE.html</a>	Natural TEs identified in <i>Drosophila melanogaster</i>
dbRIP	<a href="http://dbrip.brocku.ca/introduction.html">http://dbrip.brocku.ca/introduction.html</a>	Polymorphic TE insertions in the human genome; identification of insertions associated with genes or genomic regions

\*Jurka *et al.* (2005); †Llorens *et al.* (2011); ‡Vassetzky & Kramerov (2012).

applied to retrieve LTR retrotransposons (e.g. Flavell *et al.* 1992; Pearce *et al.* 1999) and LINEs (Alix & Heslop-Harrison 2004). Noticeably, this approach based on genomic DNA does not directly distinguish functional and active TE from degenerated and quiescent TE. Accordingly, cDNA can be used to highlight transcriptionally (and, to a certain extent, transpositionally) active TE.

With the advent of sequencing technologies, it is becoming accessible to generate genomic resources and hence highlight TE of nonmodel species. TE can be characterized *de novo* from completely sequenced and assembled genomes using a variety of existing procedures (see Makalowski *et al.* 2012; and Flutre *et al.* 2012 for comprehensive reviews). Nevertheless, reads produced by current high-throughput sequencing technologies are typically much shorter than TE whose repetitive nature hampers accurate genome assembly and downstream analyses (Treangen & Salzberg 2012). Powerful procedures exploiting data generated by superficial genome sequencing (i.e. genome snapshot with <0.1× coverage) aim at describing relatively abundant TE. Using existing databases (Table 1), the identification by blast of reads corresponding to TE sequences that are specific to a given genome is relatively straightforward for taxa that slightly diverged from model species (Wicker *et al.* 2009). Dissimilar TE remain hardly identified with such approaches, but a priori unknown TE families (e.g. from taxa lacking closely related species with well-annotated genomes) can be usefully characterized by exploiting approaches such as graph-based clustering (Novak *et al.* 2010) or assisted automated

assembly (DeBarry *et al.* 2008). In particular, the later method identifies overlapping sequences among short reads and efficiently constructs pseudomolecules representing consensus of abundant repeat families through automated genome walking (Estep *et al.* 2013). Finally, reads corresponding to individual insertions of the identified TE families can be compared, assessing patterns of genetic diversity, to infer their evolutionary dynamics and distinguish recently active from quiescent families (Senerchia *et al.* 2013). Accordingly, high-throughput sequencing offers promising approaches to identify suitable candidate TE for further examination at the population level in nonmodel species.

### Genotyping of polymorphic TE insertions in natural populations

Assessing the impact of TE on host genomes ultimately requires the screening of polymorphic insertions (i.e. TE copies at particular locations within genomes) among individuals. The most commonly used fingerprinting techniques exploit PCR to simultaneously amplify multiple TE insertions and thus detect insertional polymorphisms in TE genome fractions. These techniques are commonly referred to as transposon displays and exploit two fundamental features of repetitive TE: (i) the conserved domains among copies of a given TE family to design specific PCR primers and (ii) their transpositional activity causing structural variation across the genome to visualize distinctive insertions as fragments of different sizes. The crucial step for efficient transposon displays is

to design specific, outwards-facing primers at a conserved extremity of the candidate TE sequence. Various methods (Fig. 3), differing in the nature of the primers used and thus the sequence flanking the targeted insertions, have been developed (see Kalendar *et al.* 2011 for a recent review). Despite the lack of consistent names for these techniques, they are relatively straightforward to implement and allow surveying large numbers of individuals at reasonable costs. Transposon displays thus represent convenient strategies to explore variation in TE genome fractions in natural populations.

Inter-retrotransposon amplified polymorphism (IRAP) uses either a single primer or two primers (targeting the same or different TE families) to generate PCR fragments between nearby TEs inserted in different orientations (Kalendar *et al.* 1999). The exact same technique has been put forward for MITEs and called inter-MITE amplified polymorphism (Chang *et al.* 2001). These methods show bias towards clustered TE insertions and may be short of reproducibility as put forward for the related random amplification of polymorphic DNA (RAPD). Similar to IRAP, retrotransposon-microsatellite amplified polymorphism (REMAP) relies entirely on PCR amplifications, using a TE-specific primer and non-TE primer matching a microsatellite motif (Kalendar *et al.* 1999). It shares features, including reproducibility issues, with intersimple sequence repeat (ISSR) markers.

Sequence-specific amplified polymorphism (SSAP) is similar to amplified fragment length polymorphism (AFLP), except that it is a TE-anchored strategy. It is based on the digestion of genomic DNA with a rare and a frequent-cutter restriction enzyme to generate fragments to be PCR amplified between insertions of the candidate TE and an adaptor ligated at randomly distributed restriction sites (Syed & Flavell 2007). It thus produces fragments covering the whole genome and encompassing sequences from the termini of candidate TE insertions to their flanking genomic regions. This technique is reproducible, but may generate numerous bands and hardly scorable patterns in large genomes. Accordingly, primers with selective bases can be annealed to the adaptor site to reduce genome complexity when high-copy-number TE families are surveyed. In addition, SSAP can be easily adapted to different purposes such as the detection of variation of TE expression using cDNA (Kashkush *et al.* 2002) or changing methylation in the vicinity of TE insertions using isoschizomers (Parisod *et al.* 2009).

The above-mentioned techniques straightforwardly produce a large number of dominant loci marking insertions of candidate TEs as present or absent for further analysis with suitable approaches of population genetics (Bonin *et al.* 2007; Perez-Figueroa *et al.* 2010). Markers generated by transposon displays specifically track TE

insertions, but it should be kept in mind that the exact mechanism underlying detected polymorphisms remains hardly accessible. Assayed polymorphisms (presence/absence of bands) uncover insertional patterns associated with the transpositional activity of corresponding TEs to a large extent, but can also be due to mutations such as indels along the amplified fragments or changes of restriction sites (e.g. Petit *et al.* 2010). Comparing profiles generated by transposon displays with related fingerprinting techniques marking random sequences represents a convenient strategy to take such uncertainty into account when surveying a large number of individuals and to assess the specific dynamics of candidate TEs. For instance, AFLP and SSAP rely on very similar protocols, thus allowing to specifically track TE dynamics with SSAP while taking background genome variation into account with AFLP (Parisod & Senerchia 2013). Fragments of interest (i.e. candidate TE insertions) can also be isolated and sequenced to design locus-specific assays offering codominant markers (Devos *et al.* 2005).

Sequencing polymorphic TE insertions across genomes remains challenging, but high-throughput technologies potentially offer tremendous opportunities to be explored at the population level (Xing *et al.* 2013). In contrast to transposon displays, substantial knowledge about the candidate TEs and the host genome is necessary to implement sequencing approaches. However, they probably provide accurate descriptions of patterns and processes of variation among TE insertions and their flanking regions, offering valuable insights for evolutionary studies. So far, high-throughput sequencing of TE insertional polymorphisms has been mostly achieved in populations of cells from one or a few individuals, but hold great promises for surveying natural populations as well.

TE insertions can be sorted through bioinformatics after untargeted, whole-genome sequencing (Treangen & Salzberg 2012). In that context, sequencing platforms yielding relatively long reads offer better resolution to distinguish insertions from a given TE family and retrieve crucial information about the flanking sequence. Two main approaches have been used so far: split-read methods map and detect split reads at break points of sites of polymorphic TE insertions, whereas read-pair algorithms can make use of shorter paired-end reads to be mapped to TE sequences and flanking regions to detect structural variation indicative of polymorphic TE insertions (Xing *et al.* 2013). In particular, powerful methods have been highlighted that use reference genomes of the targeted organism, but can take a priori unknown TE insertions into account and thus estimate their frequencies at the population level (Kofler *et al.* 2012).

Methods specifically targeting candidate TE sequences before sequencing could be applied, even for



organisms without proper reference genomes. Genomic DNA can indeed be specifically enriched in fragments containing TE insertions by either PCR or sequence capture. Sequenced fragments can then be used to identify and compare flanking genomic regions among individuals at the population level. A relatively straightforward approach would be to massively sequence PCR fragments from transposon displays as otherwise achieved with AFLP (Paris & Despres 2012). Similarly, primers specific to the candidate TE families can be used together with primers annealing to the linker sequence to amplify fragments containing targeted TE insertions for the sequencing of amplicons (Witherspoon *et al.* 2010). Noticeably, PCR amplification before sequencing can be avoided by capturing fragments containing candidate TEs on custom arrays (Baillie *et al.* 2011). This approach was shown to be fairly sensitive, highlighting rare transposition events. As current high-throughput sequencing platforms yield enough data to allow for the pooling of several samples, the above-mentioned methods and forthcoming ones may be suitable for reasonably sized surveys of TE insertional patterns in natural populations of nonmodel organisms.

## Conclusions

Genomes are emerging as very dynamic entities, with their most variable fraction (i.e. TEs) probably driving changes of the overall architecture of functional genomes (Fedoroff & Bennetzen 2013). In particular, TEs seem to show bursts of activity under specific conditions that are common in the wild (Box 2) and that may translate into genome reorganization of central significance for the evolutionary ecology of the host species (Box 3). However, several interconnected processes such as selection at the TE and at the host levels may influence the evolutionary trajectories of TEs within and among taxa, and have to be better understood (Tenaillon *et al.* 2010). In addition to host-controlled transposition and TE deletion, the fate of inserted TEs is indeed determined by processes acting at the level of the host population (Le Rouzic *et al.* 2007). The distribution and accumulation of TE insertions thus depend on their effect on the host fitness and effective population size, and are thus largely determined by population features such as demography, mating system and local recombination rate that add to natural selection (Lynch 2007). Accordingly, the interplay between the environment, the host population dynamics and TE dynamics probably plays a major role in driving the evolutionary trajectories and divergence of closely related taxa. Additional work tracking TE insertions in natural populations is thus necessary to shed light on the impact of TEs (i.e. 'junk DNA') on microevolutionary processes.

## Acknowledgements

We thank A. Buerkle for his invitation to write this article and Céline Geiser as well as three anonymous reviewers for constructive comments on the manuscript. This work was supported by a Sciex-NMS<sup>ch</sup> fellowship (project N 11.235) and a grant of the Swiss National Science Foundation (PZ00P3-131950 to CP).

## References

- Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.
- Alix K, Heslop-Harrison JS (2004) The diversity of retroelements in diploid and allotetraploid *Brassica* species. *Plant Molecular Biology*, **54**, 895–909.
- Alleman M, Freeling M (1986) The Mu transposable elements of maize: evidence for transposition and copy number regulation during development. *Genetics*, **112**, 107–119.
- Baillie JK, Barnett MW, Upton KR *et al.* (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*, **479**, 534–537.
- Baucom RS, Estill JC, Chaparro C *et al.* (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genetics*, **5**, e1000732.
- Bhattacharyya M, Smith AM, Ellis THN, 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**, 115–122.
- Biémont C, Vieira C (2006) Junk DNA as an evolutionary force. *Nature*, **443**, 521–524.
- Bonin A, Ehrlich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.
- Butelli E, Licciardello C, Zhang Y *et al.* (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell*, **3**, 1242–1255.
- Capy P, Gasperi G, Biémont C, Bazin C (2000) Stress and transposable elements: co-evolution or useful parasites? *Heredity*, **85**, 101–106.
- Castel SE, Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews Genetics*, **14**, 100–112.
- Chang R-Y, O'Donoghue LS, Bureau TE (2001) Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. *Theoretical and Applied Genetics*, **102**, 773–781.
- DeBarry JD, Liu R, Bennetzen JL (2008) Discovery and assembly of repeat family pseudomolecules from sparse genomics sequence data using the assisted automated assembler of repeat families (AAARF) algorithm. *BMC Bioinformatics*, **13**, 235.
- Devos KM, Ma J, Pontaroli AC, Pratt LH, Bennetzen JL (2005) Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. *Proceedings of National Academy of Sciences USA*, **102**, 19243–19248.
- Dolgin ES, Charlesworth B (2008) The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics*, **178**, 2169–2177.
- Doolittle WF (2013) Is junk DNA bunk? A critique of ENCODE. *Proceedings of National Academy of Sciences USA*, **110**, 5294–5300.
- Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**, 601–603.
- Estep MC, De Barry JD, Bennetzen JL (2013) The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. *Heredity*, **110**, 194–204.
- Fedoroff NV (2012) Transposable elements, epigenetics, and genome evolution. *Science*, **338**, 758–767.

- Fedoroff NV, Bennetzen JL (2013) Transposons, genomic shock, and genome evolution. In: *Plant Transposons and Genome Dynamics in Evolution* (ed. Fedoroff NV), pp. 181–201. Wiley-Blackwell, Ames, IA.
- Feschotte C (2008) Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*, **9**, 397–405.
- Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) *Tyl-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acid Research*, **20**, 3639–3644.
- Flutre T, Permal E, Quesneville H (2012) Transposable element annotation in completely sequenced eukaryote genomes. In: *Plant Transposable Elements* (eds Grandbastien M-A & Casacuberta J). Topics in Current Genetics, vol. 24, pp. 17–40. Springer, Heidelberg, Germany.
- Garfinkel DJ (2005) Genome evolution mediated by Ty elements in *Saccharomyces*. *Cytogenetic and Genome Research*, **110**, 63–69.
- González J, Karasov TL, Messer PW, Petrov DA (2010) Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genetics*, **4**, e1000905.
- Grandbastien M-A (1998) Activation of plant retrotransposons under stress conditions. *Trends in Plant Science*, **3**, 181–187.
- Grandbastien M-A, Audeon C, Bonnard E *et al.* (2005) Stress activation and genomic impact of Tnt1 retrotransposons in *Solanaceae*. *Cytogenetic and Genome Research*, **110**, 229–241.
- Gray YH (2000) It takes two transposons to tango: transposable-elements-mediated chromosomal rearrangements. *Trends in Genetics*, **16**, 461–468.
- Grundt HH, Kjolner S, Borgen L, Rieseberg LH, Brochmann C (2006) High biological species diversity in the arctic flora. *Proceedings of National Academy of Sciences USA*, **103**, 972–975.
- He L, Dooner HK (2009) Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for Helitron and retrotransposon insertions. *Proceedings of National Academy of Sciences USA*, **106**, 8410–8416.
- Hollister JD, Smith LM, Guo YL *et al.* (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proceedings of National Academy of Sciences USA*, **108**, 2322–2327.
- Hu TT, Pattyn P, Bakker EG *et al.* (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, **43**, 476–481.
- Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P (2011) The struggle for life of the genome's selfish architects. *Biology Direct*, **17**, 6–19.
- Jordan IK, Rogozin IB, Glazko GV, Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics*, **19**, 68–72.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. *Cytogenetics and Genome Research*, **110**, 462–467.
- Kalendar R, Grob T, Regina M, Suoniemi A, Shulman AH (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theoretical and Applied Genetics*, **98**, 704–711.
- Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of National Academy of Sciences USA*, **97**, 6603–6607.
- Kalendar R, Flavell AJ, Ellis TH, Sjakste T, Moisy C, Schulman AH (2011) Analysis of plant diversity with retrotransposon-based molecular markers. *Heredity*, **106**, 520–530.
- Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics*, **160**, 1651–1659.
- Kidwell MG, Lisch DR (2001) Transposable elements, parasitic DNA, and genome evolution. *Evolution*, **55**, 1–24.
- Kofler R, Betancourt AJ, Schlötterer C (2012) Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genetics*, **8**, e1002487.
- de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics*, **7**, e1002384.
- Koonin EV (2009) Darwinian evolution in the light of genomics. *Nucleic Acid Research*, **37**, 1011–1034.
- Le Rouzic AL, Boutin TS, Capy P (2007) Long-term evolution of transposable elements. *Proceedings of National Academy of Sciences USA*, **104**, 19375–19380.
- Lerman DN, Feder ME (2005) Naturally occurring transposable elements disrupt Hsp70 promoter function in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **22**, 776–783.
- Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME (2003) Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. *Molecular Biology and Evolution*, **20**, 135–144.
- Li W, Zhang P, Fellers JP, Friebe B, Gill BS (2004) Sequence composition, organization, and evolution of the core Triticeae genome. *Plant Journal*, **40**, 500–511.
- Lisch D (2013) How important are transposons for plant evolution? *Nature Reviews Genetics*, **14**, 49–61.
- Liu J, He Y, Amasino R, Chen X (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes and Development*, **18**, 2873–2878.
- Llorens C, Futami R, Covelli L *et al.* (2011) The Gypsy database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acid Research*, **39**, D70–D74.
- Lynch M (2007) *The Origins of Genome Architecture*. Sinauer Associates Inc., Sunderland.
- Majewski J, Ott J (2002) Distribution and characterization of regulatory elements in the human genome. *Genome Research*, **12**, 1827–1836.
- Makalowski W, Pande A, Gotea V, Makalowska I (2012) Transposable elements and their identification. Evolutionary genomics: statistical and computational methods. Vol. 1. *Methods in Molecular Biology*, **855**, 337–359.
- Maside X, Bartolome C, Assimacopoulos S, Charlesworth B (2000) Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. *Genetics Research*, **75**, 275–284.
- Maside X, Bartolome C, Assimacopoulos S, Charlesworth B (2001) Rates of movement and distribution of transposable elements in *Drosophila melanogaster*: *in situ* hybridization vs Southern blotting data. *Genetics Research*, **78**, 121–136.
- McCart C, Ffrench-Constant RH (2008) Dissecting the insecticide-resistance-associated cytochrome P450 gene Cyp6 g1. *Pest Management Science*, **64**, 639–645.
- McClintock B (1984) The significance of responses of the genome to challenge. *Science*, **226**, 792–801.
- Melamed-Bessudo C, Levy AA (2012) Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in *Arabidopsis*. *Proceedings of the National Academy of Sciences USA*, **109**, E981–E988.
- Morgan HD, Sutherland HE, Martin DIK, Whitelaw E (1999) Epigenetic inheritance at the *agouti* locus in the mouse. *Nature Genetics*, **23**, 314–318.
- Novak P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics*, **11**, 378.
- O'Donnell KA, Burns KH (2010) Mobilizing diversity: transposable element insertions in genetic variation and disease. *Mobile DNA*, **1**, 21.
- Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. *Nature*, **284**, 604–607.
- Paris M, Despres L (2012) Identifying insecticide resistance genes in mosquito by combining AFLP genome scans and 454 pyrosequencing. *Molecular Ecology*, **21**, 1672–1686.
- Parisod C, Senerchia N (2013) Responses of transposable elements to polyploidy. In: *Plant Transposable Elements* (eds Grandbastien M-A, Casacuberta J). Topics in Current Genetics, vol. 24, pp. 147–168. Springer, Heidelberg, Germany.

- Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MA, Ainouche ML (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytologist*, **184**, 1003–1015.
- Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis THN, Flavell AJ (1999) Rapid isolation of plant *Ty1-copia* group retrotransposon LTR sequences for molecular marker studies. *Plant Journal*, **19**, 711–717.
- Perez-Figueroa A, Garcia-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.
- Petit M, Guidat C, Daniel J *et al.* (2010) Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist*, **186**, 135–147.
- Piegu B, Guyot R, Picault N *et al.* (2006) Doubling genome size without polyploidization: dynamics of retrotransposition driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Research*, **16**, 1262–1269.
- Saeidi H, Rahiminejad MR, Heslop-Harrison JS (2008) Retroelement insertional polymorphisms, diversity and phylogeography within diploid, D-genome *Aegilops tauschii* (Triticeae, Poaceae) sub-taxa in Iran. *Annals of Botany*, **101**, 855–861.
- Salvi S, Sponza G, Morgante M *et al.* (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proceedings of National Academy of Sciences USA*, **104**, 11376–11381.
- Schmidt AL, Anderson LM (2006) Repetitive DNA elements as mediators of genomic change in response to environmental cues. *Biological Reviews*, **81**, 531–543.
- Sela N, Kim E, Ast G (2010) The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome Biology*, **11**, R39.
- Seleme MC, Busseau I, Malinsky S, Bucheton A, Teninges D (1999) High-frequency retrotransposition of a marked I factor in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. *Genetics*, **151**, 761–771.
- Senerchia N, Wicker T, Felber F, Parisod C (2013) Evolutionary dynamics of retrotransposons assessed by high throughput sequencing in wild relatives of wheat. *Genome Biology and Evolution*, **5**, 1010–1020.
- Shapiro JA (2010) Mobile DNA and evolution in the 21st century. *Mobile DNA*, **1**, 4.
- Shimizu KK, Shimizu-Inatsugi R, Tsuchimatsu T, Purugganan MD (2008) Independent origins of self-compatibility in *Arabidopsis thaliana*. *Molecular Ecology*, **17**, 704–714.
- Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, **8**, 272–285.
- Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genetics*, **43**, 1160–1163.
- Syed NH, Flavell AJ (2007) Sequence-specific amplification polymorphisms (SSAPs): a multi-locus approach for analyzing transposon insertions. *Nature Protocols*, **1**, 2746–2752.
- Tenaillon MI, Hollister JD, Gaut BS (2010) A triptych of the evolution of plant transposable elements. *Trends in Plant Science*, **15**, 471–478.
- Treangen TJ, Salzberg SL (2012) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Review Genetics*, **13**, 36–46.
- Tsukahara S, Kawabe A, Kobayashi A *et al.* (2012) Centromere-targeted *de novo* integrations of an LTR retrotransposon of *Arabidopsis lyrata*. *Genes Development*, **26**, 705–713.
- Vassetzky NS, Kramerov DA (2012) SINEBase: a database and tool for SINE analysis. *Nucleic Acids Research*, **41**, D83–D89.
- Vicient CM, Jaaskelainen MJ, Kalendar R, Schulman AH (2001) Active retrotransposons are a common feature of grass genomes. *Plant Physiology*, **125**, 1283–1292.
- Vitte C, Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and Genome Research*, **110**, 91–107.
- Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. *Proceedings of National Academy of Sciences USA*, **103**, 17644–17649.
- Wang X, Weigel D, Smith LM (2013) Transposon variants and their effects on gene expression in *Arabidopsis*. *PLoS Genetics*, **9**, e1003255.
- Whitney KD, Baack EJ, Hamrick JL *et al.* (2010) A role for nonadaptive processes in plant genome size evolution? *Evolution*, **64**, 2097–2109.
- Wicker T, Sabot F, Van Hua A *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, **8**, 973–982.
- Wicker T, Taudien S, Houben A *et al.* (2009) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. *Plant Journal*, **59**, 712–722.
- Witherspoon DJ, Xing J, Zhang Y, Watkins Batzer M, Jorde LB (2010) Mobile element scanning (ME-Scan) by targeted high-throughput sequencing. *BMC Genomics*, **11**, 410.
- Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123–3135.
- Woodrow P, Pontecorvo G, Fantaccione S *et al.* (2010) Polymorphism of a new *Ty1-copia* retrotransposon in durum wheat under salt and light stresses. *Theoretical and Applied Genetics*, **121**, 311–322.
- Wright SI, Agrawal N, Bureau TE (2003) Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Research*, **13**, 1897–1903.
- Xing J, Witherspoon DJ, Jorde LB (2013) Mobile element biology: new possibilities with high-throughput sequencing. *Trends in Genetics*, **29**, 280–289.