

Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics

A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution

Christian Biémont¹

Université de Lyon, F-69000, Lyon; Université Lyon 1; Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622 Villeurbanne, France

ABSTRACT

The idea that some genetic factors are able to move around chromosomes emerged more than 60 years ago when Barbara McClintock first suggested that such elements existed and had a major role in controlling gene expression and that they also have had a major influence in reshaping genomes in evolution. It was many years, however, before the accumulation of data and theories showed that this latter revolutionary idea was correct although, understandably, it fell far short of our present view of the significant influence of what are now known as “transposable elements” in evolution. In this article, I summarize the main events that influenced my thinking about transposable elements as a young scientist and the influence and role of these specific genomic elements in evolution over subsequent years. Today, we recognize that the findings about genomic changes affected by transposable elements have considerably altered our view of the ways in which genomes evolve and work.

SINCE the radical suggestion by Barbara McClintock in the 1950s, based on her extensive genetic analyses in maize, that some genes might move along chromosomes, our knowledge of transposable elements (TEs) has vastly increased. TEs are no longer seen as “junk” and “selfish” pieces of DNA—the predominant view from the 1960s through the 1990s—but as major components of genomes that have played a significant role in evolution, an idea also first proposed by McCLINTOCK (1984: her Nobel Prize lecture). The history of these genomic elements provides one of the best examples of how scientific concepts in biology emerge and then evolve into new concepts. It is a salutary lesson for researchers, both young and old, to be tolerant of striking new ideas when they appear and not to dismiss them simply because they conflict with current theories and knowledge. History is easier to relate to when you have been a direct observer, and in this article I summarize the main events witnessed in my own scientific lifetime that paved the way to the present day understanding of the structure and composition of genomes. I make no attempt to present a complete or balanced historical account of what happened; instead,

I describe the events and discoveries that influenced the thinking of a young scientist, not just despite, but because of, their very strangeness and incompatibility with what was then received opinion.

TRANSPOSABLE ELEMENTS AS COMPONENTS OF GENETIC DIVERSITY

In the 1970s, the field of population genetics was dominated by analysis of the genetic polymorphism of populations using allozymes, with the aim of deciphering population structuring (LEWONTIN 1974). The entire emphasis was on the role of point mutations in coding regions as the primary source of evolutionary change. Despite the observations of McCLINTOCK (1950) that in maize some genetic factors [*e.g.*, the Activator (*Ac*)/Dissociator (*Ds*)] that can control the cell color of kernels were able to change their locations within and between chromosomes and could control the expression of some genes [see FEDOROFF (1994) for a biography of B. McCLINTOCK], and the demonstrated presence of mobile DNA elements in bacteria (SHAPIRO 1969), the possibility that TEs could influence genetic polymorphism, and therefore genetic diversity, was, for the most part, ignored. Some researchers did not even believe that these unconventional DNA insertions

¹Address for correspondence: UMR 5558, Biométrie et Biologie Evolutive CNRS, Université Lyon1, 69622 Villeurbanne Cedex, France.
E-mail: christian.biémont@univ-lyon1.fr

actually moved or moved with significant frequency and so thought that they could not possibly contribute to genetic diversity. This was because they always found the same TE insertion in the particular mutant in which they were interested. Such a result was, however, to be expected, given the screening process that scientists were using to isolate their mutant.

However, the 1970s saw the emergence in *Drosophila* research of the hybrid dysgenesis phenomenon, in which crosses between specific lines of *Drosophila melanogaster* led to various genetic changes, including sterility and increased mutation and recombination rates. It was some years before these effects were finally shown to be associated with the mobilization of specific TEs: *P* elements (for the *P/M* system) and *I* elements (for the *I/R* system) (see PICARD *et al.* 1978; KIDWELL 1979; RUBIN *et al.* 1982; ENGEL 1988). *P* elements were later shown to be DNA transposons (TEs that transpose via a DNA intermediary) and *I* elements to be non-long terminal repeat (LTR) retrotransposons [TEs that have no long terminal repeat at their extremities and that transpose via an RNA intermediary (FINNEGAN 1992)]. Yet even this simple division does not do justice to the known complexity of the entire set of TEs, as revealed over the past 30 years. Some of these new TEs, for example, have a composite structure and appear to be major components of some genomes. The LTR retrotransposons, which have a long terminal repeat at their extremities, have been divided into various subfamilies: Ty1-copia-like (Pseudoviridae), Ty3-gypsy-like (Metaviridae), and Pao-BEL-like, depending on their sequence similarity and the order of the gene products that they encode. Because Barbara McClintock's *Ac* elements in maize and the *P* elements in *Drosophila*, which both transpose via a DNA intermediary, were initially named "transposons," it became customary to use this term as a generic term for all kinds of transposable elements, regardless of whether their mechanism of transposition involved DNA or RNA. To avoid confusion while still taking historical usage into account, it is now usual to use the term "DNA transposons" for the DNA-based TEs and the term "retrotransposons" (either with or without a long terminal repeat at their extremities) for the RNA-based TEs. To bring some order to the nomenclature in the field, WICKER *et al.* (2007) and KAPITONOV and JURKA (2008) have proposed a new TE classification based on transposition mechanism, sequence similarities, and structural relationships so as to include the new TE classes that have emerged.

The discoveries that *P* and *I* elements were actually TEs was decisive in convincing us that TEs were of interest, not only because we realized that they were able to move within the genome, but also because the *P* element had been shown to have invaded all known populations of *D. melanogaster* worldwide within the space of ~50 years (ANXOLABÉHÈRE *et al.* 1988), after

having been transferred into this species from *Drosophila willistoni* (DANIELS *et al.* 1990). Fifty years is a very short period in terms of evolutionary time, but a manageable interval for population geneticists. The apparently deleterious effects of the mobilization of the *P* and *I* elements as a result of crosses, however, confirmed the idea that TEs are "selfish"; *i.e.*, they produce only detrimental effects on organisms, an idea developed in 1980 by DOOLITTLE and SAPIENZA, which was generally accepted for many years. Were this simple idea true, the evolutionary significance of TEs would be slight. But during the 1980s, new findings were already showing the picture to be more complex.

In particular, various TEs were discovered by molecular biologists who were interested either in the composition of the genomes or in the sequences of mutant alleles. Several Russian teams were in the vanguard of this research, and they referred to the first TEs of *Drosophila* as "mobile dispersed genes" (*mdg*), a term that everybody could understand. *mdg1*, *mdg2*, *mdg3*, and *mdg4* were the first to be discovered and described (ANANIEV and ILYIN 1981; KULGUSKIN *et al.* 1981; GVOZDEV *et al.* 1981). This numerical nomenclature was soon superseded by more colorful names such as *gypsy* (*mdg4*), *copia*, and 412. Since then, many other names have emerged from the imaginative minds of scientists in a seemingly unending flow and, in a way, it is a pity that these imaginative names will finally have to be replaced by a more austere but accurate and much-needed systematic nomenclature reform, as noted above (FINNEGAN 1992; WICKER *et al.* 2007; KAPITONOV and JURKA 2008).

The next phase in the exploration of these elements was to estimate their number of copies by Southern blots and to localize precisely their chromosomal insertions, as could be visualized on the polytene chromosomes of *Drosophila*, using *in situ* hybridization. Initially, such localization was carried out with radioactive labels but then with the more powerful biochemical labeling technique based on avidin-streptavidin and dyes. At the end of the 1970s, the period when I was a post doc student in the United States, I first came into contact with the *in situ* hybridization technique and some "strange things" about TEs. In particular, I remember being shown some pictures of *Drosophila* salivary gland polytene chromosomes taken by a Ph.D. student, which showed strong labeling in a specific band in one chromosome but no labeling on its homolog (the two chromosomes had been separated on the squash as sometimes happens). This young researcher, using a probe of a specific TE, showed me these pictures surreptitiously because it was difficult at this time to understand what this was all about. Furthermore, the *in situ* hybridization technique led to estimations of the number of copies of various TEs and detection of the polymorphism of their insertion in different strains and populations. Such polymorphism provided direct evidence of the continuing mobility of these elements.

Most of these studies were done on *D. melanogaster*, which has a low frequency of TE insertions at numerous, albeit limited, numbers of sites (usually less than 100 in this species). One surprising result was the observation of transposition bursts of TEs such as Doc (a non-LTR retrotransposon; GERASIMOVA *et al.* 1990), copia (a retrotransposon; BIÉMONT *et al.* 1987; GERASIMOVA *et al.* 1990; PASYUKOVA and NUZHIDIN 1993), and *P* (a DNA transposon; BIÉMONT *et al.* 1990) while inbred lines were being maintained in the laboratory. Although we did not know whether this mobilization was the result of the inbreeding itself, or simply “spontaneous” transposition revealed by the homozygosity of the lines, which made it easier to observe new insertions, we were forced to conclude that inbred lines were not as homozygous as expected, at least with regard to TE insertions. This opened the way for the further observation that TE movement and the resulting polymorphism could influence quantitative traits during artificial selection (SHRIMPTON *et al.* 1990). While empirical data accumulated about various strains and populations (YAMAGUCHI *et al.* 1987; CHARLESWORTH *et al.* 1992; BIÉMONT *et al.* 1994), theoretical approaches and simulations—the parameters of which involved transposition and excision rates, selection against the insertion of TEs, and effective population size—furnished useful models with which to confront the experimental data with some theoretical expectations (CHARLESWORTH and CHARLESWORTH 1983).

These population-based models now look imperfect and oversimplified, mainly because they analyzed TE dynamics in terms of asymptotic values, with little understanding of the dynamics of the variation in TE copy number over long stretches of time. Nevertheless, at the time, these models were very useful in forcing experimental scientists to reanalyze their data, to estimate and reconsider the parameters that they were using in the models, and to invent new protocols. It is striking to note, however, that the precise mechanisms underlying the action of natural selection against TE insertions, which involve either gross chromosomal rearrangements caused by unequal recombination between TE copies [the ectopic exchange model of LANGLEY *et al.* (1988)] or the slightly deleterious effects of TE insertions that reduce host fitness, are still topics of considerable debate (CHARLESWORTH *et al.* 1997; BIÉMONT *et al.* 1997), with most data now coming from plants (TIAN *et al.* 2009; LOCKTON and GAUT 2010) and humans (SONG and BOISSINOT 2006), in which the reproductive system (HICKEY 1982; DOLGIN *et al.* 2008), demography (LOCKTON *et al.* 2008), and population size (LYNCH and CONERY 2003) seem to play major roles. In the present context, it is interesting to remember some discussions in which the possibility of the presence or influence of TEs in organisms “higher” than bacteria or *Drosophila*, such as humans or plants, was dismissed out of hand. However, DNA-reassociation studies had al-

ready shown us long before that genomes of many organisms, including plants, do in fact contain many repeated sequences (FLAVELL *et al.* 1974; BALDARI and AMALDI 1976; CRAIN *et al.* 1976; VENTURINI *et al.* 1987). The connection between such sequences and TEs would not be demonstrated for many years and, as a result, we found it hard to accept that the human genome is in fact full of TEs and other repeated sequences, a point that was definitively admitted only when the human genome had been fully sequenced. Plant genomes were also initially assumed to be free of retrotransposons until these elements were actually looked for, and we now know that the genomes of some plants contain large numbers of TEs of this kind (VOYTAS and AUSUBEL 1988) (see Figure 1 for the proportion of TEs in various organisms). Most genomes appear to contain a mixture of TEs, some of which are still active while others are ancient relics that have degenerated and are sometimes no longer recognizable as TEs. Because we can now detect the presence of ancient TE copies or decayed sequences that are the hallmarks of TEs, the proportion of the genome now known to have originated from TEs is increasing.

One interesting debate that began in earnest in the 1990s concerned the relationship between retroviruses and retrotransposable elements. Some people thought that TEs arose from retroviruses, but made no attempt to explain where the retroviruses had come from, a view that, like the hypothesis of the extraterrestrial origin of life, is not *per se* unlikely, but in the end only shifts the question elsewhere. Although it was easier to accept that retroviruses and the members of the LTR retrotransposon family had common ancestors, and that retroviruses arose from TEs by the addition of an envelope gene, even these ideas were not very readily accepted (VARMUS 1988). This may have been because TEs did not have the status that viruses had then and because it was easier to recognize and admit the existence of viruses, with which we were used to coexisting, than to accept the idea of the presence of thousands of copies of junk DNA actually inside our human genome. The discovery that some TEs were capable of producing virus-like particles similar to those of retroviruses (MOSSIE *et al.* 1985; MIYAKE *et al.* 1987), and the similarity of their reverse transcriptase-like sequences (XIONG and EICKBUSH 1988), clearly revealed the link between retrotransposons and retroviruses (see VARMUS 1988). It was also at this time that infectious retroviruses were discovered in *Drosophila* (KIM *et al.* 1994; SONG *et al.* 1994), making this species once again a good model for the analysis of the dynamics of such elements within genomes and populations and between species. Such infectious retroviruses were subsequently also identified in plants (WRIGHT and VOYTAS 1998). Because infectious retroviruses can be transmitted between organisms, this greatly broadened the notion of the possible horizontal transfer of TEs, one example of which had been

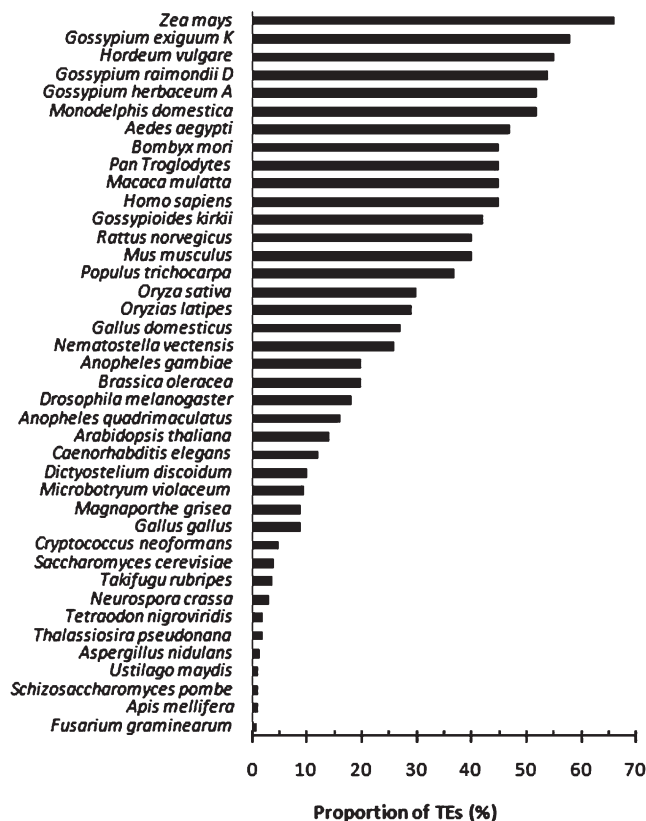


FIGURE 1.—Estimated proportion of TEs in genomes of various eukaryotic organisms. The data consist only of the TEs clearly identified as such. Parts of the genome that had simply been derived from TEs—such as TE-derived non-TE sequences (*i.e.*, sequences derived from TEs, but no longer recognizable as TEs and that appear as “unique” DNA sequences), non-TE sequences duplicated as a result of TE activity, TE-derived sequences included within genes, satellite DNA, and minisatellites and microsatellites derived from TEs (PAVLÍČEK *et al.* 2002; JURKA and GENTLES 2006; MA and JACKSON 2006)—have not been taken into account. Courtesy of C. Lævenbruck.

unambiguously established for the transfer of the *P* DNA transposon between two species of *Drosophila*.

One important clue discovered toward the end of the 1980s was the observation that some TEs contain internal sequences similar to those found in murine mammary tumor virus LTRs, which are known to play a role in the regulation of the provirus by steroid hormones (PERONNET *et al.* 1986; ZIARCZYK and BEST-BELPOMME 1991), and element 412 was found to contain a 20-HD ecdysone-responsive repetitive sequence (MICARD *et al.* 1988). In addition to such hormone-sensitive sequences, the presence of sequences homologous to heat-shock consensus sequences suggested the possibility that TEs could be sensitive to their environment. It is surprising that such observations have not been investigated further because hormones are known to regulate various genes during development, and their action on TEs could help to include TEs in gene network regulatory systems (see below).

It was during the 1990s that retroviruses and TEs began to be used as vectors to transfer genes within or between species and therefore became important tools in genetic engineering. While many experiments used retroviruses to insert new genes, especially in higher organisms, many experiments also tried to use LTR retrotransposons and DNA transposons with the aim of obtaining a powerful universal vector. We are still far from achieving this objective, although recent discoveries look promising (GRABUNDZIJA *et al.* 2010).

TRANSPOSABLE ELEMENTS AS PLAYERS IN EVOLUTION

Except for their use as potential genetic tools, the interest in molecular analyses of TEs faded between 1990 and 2000, and the population approach was not really understood. Researchers were interested in the precise mechanisms by which TE activity and copy number are regulated, rather than in global processes such as those involved at the level of populations. Some forces do indeed select against TE insertions (due to the deleterious impact of insertions or of their effects through ectopic recombination), and drift, resulting from the small effective population size of the host, may be at work even for TEs that are strongly regulated at the molecular level. However, there was a great renewal of interest when sequenced genomes of organisms such as humans, mice, and plants became available. The large amount of TEs present in these organisms forced us to finally reconsider the assumption that the TEs were purely selfish and to envisage instead that they (or some of their insertions) may have evolved toward genomic functions (they are said to have been “domesticated”) (MCDONALD 1983). This idea of a possible genomic function of some TE insertions was initially rejected, mostly because no fixed TE insertions had been identified in chromosomes. However, this was mainly due to the *in situ* hybridization technique used, which tended to detect large insertions rather than smaller ones, such as the solo LTRs of retrotransposons that were later shown to be associated with gene regulation. In addition to the accepted notion that specific TE domestication has led to the RAG genes of the immunoglobulin system, which are derived from DNA transposons and act like a transposase in the V(D)J recombination system (KAPITONOV and JURKA 2005), and the use of the TART and HeT elements as protectors of telomeres in *Drosophila* (PARDUE and DEBARYSHE 2003) instead of the classical sequences usually found in many organisms, or the acquisition of novel cellular functions by recruitment of TE-derived coding sections (MILLER *et al.* 1999), the TEs were also seen as responsible for gene regulation.

On the basis of many recent findings, we now consider that TEs have considerably shaped the structure, function, and evolution of the genomes and that

the regulatory sequences that they possess can interfere with the networks of regulation of many genes, even of genes located at some distance from them (FESCHOTTE 2008; HERPIN *et al.* 2010). TEs must therefore be considered to be integrated components of the genomes, which have played a major role in evolution. It is proposed that it might be more accurate to view the genome as an ecological ecosystem in which the TE families and subfamilies correspond to the “species” of an ecosystem (BROOKFIELD 2005; MAURICIO 2005; LE ROUZIC *et al.* 2007; VENNEN *et al.* 2009). These ideas suggest that the term “controlling element,” initially proposed by B. McClintock, was appropriate, even if all TE insertions do not have this “gene-controlling” capacity. In the 1980s, I read in a book that B. McClintock had been right except in her suggestion that TEs could act as “controlling elements.” This illustrates the difficulty of changing one’s views, perhaps especially when our own genome is concerned. The idea that TEs could control some genes appeared more reasonable when some insertions of even a part of their sequence, such as a LTR that possesses promoters, were found in permanent positions with respect to specific genes. However, this does not eliminate the possibility that some more recent TE insertion sites may also be involved in gene network regulation and in species adaptation (GONZÁLEZ *et al.* 2010). The influence that TEs have had and are still having on genomes is not diminished even if the effects of some of their insertions are neutral and detrimental and contribute to the inbreeding genetic load or display selfish behavior. It is also gratifying to realize that the existence of TEs accounts for most of the early reports of “spontaneous” mutations observed in some natural populations or laboratory lines, which suggested the presence of “mutator genes or mutator systems” (DEMEREK 1927; DUSEEVA 1948; TINYAKOV 1939; GREEN 1973).

The mutational capacity of the TEs, their power to regulate genetic systems, and their sensitivity to environmental stress that has been shown to mobilize them, soon led to the idea that the TEs not only could generate genetic polymorphism favoring population adaptation, but also could promote speciation (FLAVELL 1982; GEORGIEV *et al.* 1983; McDONALD 1983; SYVANEN 1984). Because the effects of TEs were seen at that time as being predominantly deleterious, some authors concluded that TEs would have only the effect of increasing the genetic load, rather than leading to speciation (KRIEBER and ROSE 1986). We now have considerable evidence showing the impact of the TEs throughout evolution (BIÉMONT and VIEIRA 2006; FESCHOTTE and PRITHAM 2007), possibly through chromosomal rearrangements (KING 1995; see ZHANG *et al.* 2009 in maize), but the possibility that TEs may have facilitated or promoted speciation is still the subject of considerable debate (FONTDEVILA 2005;

REBOLLO *et al.* 2010). This is because it is difficult to find out whether a change in TE content or activity during a specific evolutionary period is the cause or the consequence of the speciation process.

In particular, we are not yet able to link TE characteristics, such as a high copy number, high transposition or transcription activity to the ability of populations or species to adapt better to new environmental conditions. Is there any relationship between evolutionary radiation and the composition of the genomes in terms of TEs and other repetitive sequences? Does *Bombyx mori*, with its 45% of TEs (INTERNATIONAL SILKWORM GENOME CONSORTIUM 2008), actually do any better or have stronger “evolvability,” than *Apis mellifera* (HONEYBEE GENOME SEQUENCING CONSORTIUM 2006) with its puny 1% of TEs? And what is the biological significance when a plant or an amphibian genome is composed of >70% TEs? Does the increase in the TE amount seen in some invasive species result from confronting fresh environmental conditions (VIEIRA *et al.* 1999)? Do the genomes really need TEs? More comparative analyses of genomes of various organisms and various populations of the same species will be required before we can answer these questions. This will be facilitated by improvements in sequencing techniques that will allow the low-cost processing of data from many individuals and many species. This growing body of genomic sequence data will provide precise information about the exact localization and precise DNA sequence of the copies of all kinds of TE families in many individuals, populations, and species in various ecological and physiological environments, allowing the history of genome invasion to be inferred and the impact of the environment to be evaluated. We hope, however, that the sequences of the TEs and other repeated sequences will also be taken into consideration in these studies and that they will not be ignored as has happened all too often because their annotation is difficult and time-consuming. These large-scale DNA sequence analyses should make it possible to identify new TE families, thus helping to explain why only some TE families have invaded genomes efficiently, and why some TE families have invaded certain genomes but not others. In addition, there should be real progress in the analysis of epigenetic marks, which have been known since the 1980s to interfere with TE activity (SOBIESKI and EDEN 1981). It is now clear that DNA methylation of cytosines, histone modifications, and RNA interference, all interdependent mechanisms associated with chromatin conformation, can switch TEs on or off. These processes must therefore play a role not only in defending the genome against invasion by TEs and retroviruses, but also in the complex interactions involved in gene regulation throughout development (HUDA *et al.* 2010), as reported for neuronal development, as well as in developmental processes in mouse oocytes and preimplantation embryos (MUOTRI

et al. 2007; SASAKI *et al.* 2008). The possibility remains, however, that the presence of the TEs in many eukaryote genomes is also the result of their having been selected as components of (inactive) heterochromatin because of their importance in chromosomal biology and cell division (BIÉMONT 2009). We must not forget the role of TEs in the spread of antibiotic resistance genes in bacterial populations, which can have a dramatic impact on human health, and their role via their epigenetic activation-inactivation in the effects of the early environment (nutrition, ultraviolet light, temperature) to adulthood (BOLLATI and BACCARELLI 2010), making them significant players in the interaction between genotype and phenotype.

New models of TE copy-number dynamics will be developed on the basis of the new epigenetic regulations that involve RNA interference (LU and CLARK 2010). While such models will be more precise and will have to be applied to each specific TE family, they will not entirely replace the more conventional models based on the transposition-excision rate, selection coefficient, and effective population size. These models still have their utility at the level of populations, especially because population structuring may have a greater impact on TE dynamics than previously imagined (DECÉLIERE *et al.* 2005). The numbers of TE sequences to be analyzed will be large, and larger than that obtained from *in situ* hybridizations and Southern blots (STROBEL *et al.* 1979; CHARLESWORTH *et al.* 1992; BIÉMONT *et al.* 1994), but we must be aware that while new questions will emerge from these data, the old questions, such as the analysis and modeling of the dynamics of TE families, will still depend on an appropriate definition of what a TE family is. Because a TE family is defined as a “set of phylogenetically close copies that share >80% sequence identity” (WICKER *et al.* 2007), the new data will not be very different from the data obtained by *in situ* hybridization, which specifically detected and identified sequences on the basis of such similarity criteria (CHARLESWORTH *et al.* 1992; BIÉMONT *et al.* 1994).

As the result of the ongoing sequencing of the genomes of a wide variety of organisms, we will be faced with the need to analyze huge amounts of data, and will enter a cycle of complexity in which the genome will be considered in its entirety. TEs and all other repeated sequences of the genome could play a role in deciphering this complexity (SHAPIRO 2010). While we will inevitably tend to concentrate on the human genome and on the genomes of organisms of industrial or agronomic interest, we should not overlook the historical model organisms, such as *Drosophila*, yeast, nematodes, and *Arabidopsis*, because the knowledge that we have accumulated about them over decades of study constitutes a mine of information that will enable us to develop new ideas and approaches to understanding our own genome.

CONCLUDING REMARKS

TEs were initially considered to be just junk DNA, with no influence on the genes and endowed solely with the capacity to invade genomes and populations because of their ability to transpose. They began to acquire their more exalted status once we recognized their influence on recombination rates and chromosomal rearrangements, their role as mutators and gene regulators, and their ability to be domesticated by the genome, which transforms them into new genes (VOLFF 2006; SINZELLE *et al.* 2009). Thirty years ago we began to suspect that these repeated sequences could possibly have some influence on genetic diversity, but it was beyond even the most fertile minds to imagine all their properties and their huge impact on genome composition and genome functioning. So far molecular biologists have concentrated mainly on the 1–2% of the genome composed of protein-coding genes; we now have to incorporate all the sequences that surround these genes if we are to have an overall view of all the forces that allow genomes to define an organism in interaction with the environment. In addition to incorporating TEs into genetic networks (RAMOS *et al.* 2007; FESCHOTTE 2008), we will have to understand how some repeated sequence–gene interactions can be “redundant,” thus enabling genomes to cope with huge changes in their structure and composition. We will then have to consider all the repeated and nonrepeated sequences. Some sequences will be found to play major roles while others will undoubtedly be found to be only selfish. Such research is likely to have a major impact in the field of cancer, some of the phenomenology of which involves both epigenetic processes and TE reactivation (SERAFINO *et al.* 2009; LAMPRECHT *et al.* 2010; SHALGI *et al.* 2010; WILKINS 2010). TEs and all the other repeated sequences might, once again, have some major new surprises in store for us.

It is very satisfying for a population geneticist to see that whereas at the end of the 1970s we wondered whether TEs really were involved in the generation of biologically significant genetic variability, and whether there was any point in analyzing them in natural populations, today the study of variation at the genome level has settled the question. This is due to the discovery of numerous SNPs in the genomes of different individuals and of the great variation in epigenetic marks observed between individuals, tissues, and even cells (BIÉMONT 2010; JOHNSON and TRICKER 2010; MELCER and MESHORER 2010; SKIPPER *et al.* 2010). The history of transposable elements is a good example of how science works and of how new concepts can be progressively incorporated and evolved until, in the end, they entirely transform our way of looking at things. In the 1950s, Barbara McClintock’s ideas that some DNA sequences were able to move between different sites in the chromosomes, and that some were

involved in the control of gene expression, were difficult to understand and accept, but she has been fully vindicated. Her thinking represented a major step forward in our understanding of how genomes work. Although we are still far from a clear or complete vision of gene network interactions and regulation, major progress is being made in the analysis of complex systems. In this research, TEs and epigenetics and epigenomics can be expected to play a major role, a far larger one than could have been anticipated even 10 years ago.

I thank my colleagues E. Lerat and C. Vieira for their helpful comments, A. S. Wilkins for his useful suggestions that greatly improved the text, and Monika Ghosh for her invaluable help with the English language during all these years.

LITERATURE CITED

- ANANIEV, E. V., and Y. V. ILYIN, 1981 A comparative study of the location of mobile dispersed genes in salivary gland and midgut polytene chromosomes of *Drosophila melanogaster*. *Chromosoma* **82**: 429–435.
- ANXOLABÉHÈRE, D., M. G. KIDWELL and G. PERIQUET, 1988 Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P elements. *Mol. Biol. Evol.* **5**: 252–269.
- BALDARI, C. T., and F. AMALDI, 1976 DNA reassociation kinetics in relation to genome size in four amphibian species. *Chromosoma* **59**: 13–22.
- BIÉMONT, C., 2009 Are transposable elements only silenced or are they under house arrest? *Trends Genet.* **25**: 333–334.
- BIÉMONT, C., 2010 From genotype to phenotype. What do epigenomics and epigenetics tell us? *Heredity* **105**: 1–3.
- BIÉMONT, C., and C. VIEIRA, 2006 Junk DNA as an evolutionary force. *Nature* **443**: 521–524.
- BIÉMONT, C., A. AOUAR and C. ARNAULT, 1987 Genome reshuffling of the copia element in a *Drosophila melanogaster* inbred line. *Nature* **329**: 742–744.
- BIÉMONT, C., C. ARNAULT, A. HEIZMANN and S. RONSSERAY, 1990 Massive changes in genomic locations of P elements in an inbred line of *Drosophila melanogaster*. *Naturwissenschaften* **77**: 485–488.
- BIÉMONT, C., F. LEMEUNIER, M. GARCIA GUERREIRO, J. F. BROOKFIELD, C. GAUTIER *et al.*, 1994 Population dynamics of the *copia*, *mdg1*, *mdg3*, *gypsy* and P transposable elements in a natural population of *Drosophila melanogaster*. *Genet. Res.* **63**: 197–212.
- BOLLATI, V., and A. BACCARELLI, 2010 Environmental epigenetics. *Heredity* **105**: 105–112.
- BROOKFIELD, J. F. Y., 2005 The ecology of the genome: mobile DNA elements and their hosts. *Nat. Rev. Genet.* **6**: 128–136.
- CHARLESWORTH, B., and D. CHARLESWORTH, 1983 The population dynamics of transposable elements. *Genet. Res.* **42**: 1–27.
- CHARLESWORTH, B., A. LAPID and D. CANADA, 1992 The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genet. Res.* **60**: 103–114.
- CHARLESWORTH, B., C. H. LANGLEY and P. SNIEGOWSKI, 1997 Transposable element distribution in *Drosophila*. *Genetics* **147**: 1993–1995.
- CRAIN, W. R., E. H. DAVIDSON and R. J. BRITTEN, 1976 Contrasting patterns of DNA sequence arrangement in *Apis mellifera* (honeybee) and *Musca domestica* (housefly). *Chromosoma* **59**: 1–12.
- DANIELS, S. B., K. R. PETERSON, L. D. STRAUSBAUGH, M. G. KIDWELL and A. CHOVIK, 1990 Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* **124**: 339–355.
- DECELIÈRE, G., S. CHARLES, C. VIEIRA and C. BIÉMONT, 2005 The dynamics of transposable elements in structured populations. *Genetics* **169**: 467–474.
- DEMEREK, M., 1927 The behavior of mutable gene. *Proceedings of the Fifth International Congress of Genetics Inheritance*, Berlin, pp. 183–193.
- DOLGIN, E. S., B. CHARLESWORTH and A. D. CUTTER, 2008 Population frequencies of transposable elements in selfing and outcrossing *Caenorhabditis nematodes*. *Genet. Res.* **90**: 317–329.
- DOOLITTLE, W. F., and C. SAPIENZA, 1980 Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**: 601–603.
- DUSEEVA, N. D., 1948 On the distribution of high mutability in populations of *Drosophila melanogaster*. *Dokl. Akad. Nauk SSSR* **59**: 151–159.
- ENGEL, W. R., 1988 P elements in *Drosophila*, pp. 437–484 in *Mobile DNA*, edited by D. E. BERG and M. HOWE. ASM publications, Washington, DC.
- FEDOROFF, N. V., 1994 Barbara McClintock (June 16, 1902–September 2, 1992). *Genetics* **136**: 1–10.
- FESCHOTTE, C., 2008 Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**: 397–405.
- FESCHOTTE, C., and E. J. PRITHAM, 2007 DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**: 331–368.
- FINNEGAN, D. J., 1992 Transposable elements. *Curr. Opin. Genet. Dev.* **2**: 861–867.
- FLAVELL, R., 1982 Sequence amplification, deletion and rearrangement: major sources of variation during species divergence, pp. 301–323 in *Genome Evolution*, edited by G. DOVER and R. FLAVELL. Academic Press, London.
- FLAVELL, R. B., M. D. BENNETT, J. B. SMITH and D. B. SMITH, 1974 Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* **12**: 257–269.
- FONTDEVILA, A., 2005 Hybrid genome evolution by transposition. *Cytogenet. Genome Res.* **110**: 49–55.
- GEORGIEV, G. P., D. A. KRAMEROV, A. P. RYSKOV, K. G. SKRYABIN and E. M. LUKANIDIN, 1983 Dispersed repetitive sequences in eukaryotic genomes and their possible biological significance. *Cold Spring Harb. Symp. Quant. Biol.* **47**: 1109–1121.
- GERASIMOVA, T. G., L. A. BOLENIKOVA, S. L. KISELEV and N. A. TCHURIKOV, 1990 Participation of new mobile elements in transposition bursts in *Drosophila melanogaster*. *Dokl. Akad. Nauk SSSR* **311**: 474–476.
- GONZÁLEZ, J., T. L. KARASOV, P. W. MESSER and D. A. PETROV, 2010 Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genet.* **6**: e1000905.
- GRABUNDZIJA, I., M. IRGANG, L. MÁTÉS, E. BELAY, J. MATRAI *et al.*, 2010 Comparative analysis of transposable element vector systems in human cells. *Mol. Ther.* **10**: 1200–1209.
- GREEN, M. M., 1973 Some observations and comments on mutable and mutator genes in *Drosophila*. *Genetics* **73**: 187–194.
- GVOZDEV, V. A., E. S. BELYAEVA, Y. V. ILYIN, I. S. AMOSOVA and L. Z. KAIDANOV, 1981 Selection and transposition of mobile dispersed genes in *Drosophila melanogaster*. *Cold Spring Harb. Symp. Quant. Biol.* **45** (Pt. 2): 673–685.
- HERPIN, A., I. BRAASCH, M. KRAEUSLING, C. SCHMIDT, E. C. THOMA *et al.*, 2010 Transcriptional rewiring of the sex determining *dmrt1* gene duplicate by transposable elements. *PLoS Genet.* **6** (2): e1000844.
- HICKEY, D. A., 1982 Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101**: 519–531.
- HONEYBEE GENOME SEQUENCING CONSORTIUM, 2006 Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931–949.
- HUDA, A., L. MARINO-RAMÍREZ and I. K. JORDAN, 2010 Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. *Mob. DNA* **1**: 2.
- INTERNATIONAL SILKWORM GENOME CONSORTIUM, 2008 The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem. Mol. Biol.* **38**: 1036–1045.
- JOHNSON, L. J., and P. J. TRICKER, 2010 Epigenomic plasticity within populations: its evolutionary significance and potential. *Heredity* **105**: 113–121.
- JURKA, J., and A. J. GENTLES, 2006 Origin and diversification of minisatellites derived from human Alu sequences. *Gene* **365**: 21–26.
- KAPITONOV, V. V., and J. JURKA, 2005 RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* **3** (6): e181.

- KAPITONOV, V. V., and J. JURKA, 2008 A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* **9**: 411–412.
- KIDWELL, M. G., 1979 Hybrid dysgenesis in *Drosophila melanogaster*: the relationship between the P-M and I-R interaction systems. *Genet. Res.* **33**: 205–217.
- KIM, A., C. TERZIAN, P. SANTAMARIA, A. PÉLISSON, N. PRUD'HOMME *et al.*, 1994 Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **91**: 1285–1289.
- KING, M., 1995 Species evolution: the role of chromosome change. *Syst. Biol.* **44**: 578–580.
- KRIEBER, M., and M. R. ROSE, 1986 Molecular aspects of the species barriers. *Annu. Rev. Ecol. Syst.* **17**: 465–485.
- KULGUSKIN, V. V., Y. V. ILYIN and G. P. GEORGIEV, 1981 Mobile dispersed genetic element MDG1 of *Drosophila melanogaster*: nucleotide sequence of long terminal repeats. *Nucleic Acids Res.* **9**: 3451–3464.
- LAMPRECHT, B., K. WALTER, S. KREHER, R. KUMAR, M. HUMMEL *et al.*, 2010 Derepression of an endogenous long terminal repeat activates the CSF1R protooncogene in human lymphoma. *Nat. Med.* **16**: 571–579.
- LANGLEY, C. H., E. A. MONTGOMERY, R. HUDSON, N. KAPLAN and B. CHARLESWORTH, 1988 On the role of unequal exchange in the containment of transposable element copy number. *Genet. Res.* **52**: 223–235.
- LE ROUZIC, A., S. DUPAS and P. CAPY, 2007 Genome ecosystem and transposable elements species. *Gene* **390**: 214–220.
- LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.
- LOCKTON, S., and B. S. GAUT, 2010 The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. *BMC Evol. Biol.* **10**: 10.
- LOCKTON, S., J. ROSS-IBARRA and B. S. GAUT, 2008 Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proc. Natl. Acad. Sci. USA* **105**: 13965–13970.
- LU, J., and A. G. CLARK, 2010 Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Res.* **20**: 212–227.
- LYNCH, M., and J. S. CONERY, 2003 The origins of genome complexity. *Science* **302**: 1401–1404.
- MA, J., and S. A. JACKSON, 2006 Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. *Genome Res.* **16**: 251–259.
- MAURICIO, R., 2005 Can ecology help genomics: The genome as ecosystem? *Genetica* **123**: 205–209.
- MCCLEINTOCK, B., 1950 The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **36**: 344–355.
- MCCLEINTOCK, B., 1984 The significance of responses of the genome to challenge. *Science* **226**: 792–801.
- MCDONALD, J. F., 1983 The molecular basis of adaptation: a critical review of relevant ideas and observations. *Annu. Rev. Ecol. Syst.* **14**: 77–102.
- MELCHER, S., and E. MESHORER, 2010 Chromatin plasticity in pluripotent cells. *Essays Biochem.* **48**: 245–262.
- MICARD, D., J. L. COUDERC, M. L. SOBRIER, G. GIRAUD and B. DASTUGUE, 1988 Molecular study of the retrovirus-like transposable element 412, a 20-OH ecdysone responsive repetitive sequence in *Drosophila* cultured cells. *Nucleic Acids Res.* **16**: 455–470.
- MILLER, W. J., J. F. McDONALD, D. NOUAUD and D. ANXOLABÉHÈRE, 1999 Molecular domestication: more than a sporadic episode in evolution. *Genetica* **107**: 197–207.
- MIYAKE, T., N. MAE, T. SHIBA and S. KONDO, 1987 Production of virus-like particles by the transposable genetic element, copia, of *Drosophila melanogaster*. *Mol. Gen. Genet.* **207**: 29–37.
- MOSSIE, K. G., M. W. YOUNG and H. E. VARMUS, 1985 Extrachromosomal DNA forms of copia-like transposable elements, F elements and middle repetitive DNA sequences in *Drosophila melanogaster*: Variation in cultured cells and embryos. *J. Mol. Biol.* **182**: 31–43.
- MUOTRI, A. R., M. C. MARCHETTO, N. G. COUFAL and F. H. GAGE, 2007 The necessary junk: new functions for transposable elements. *Hum. Mol. Genet.* **16**: 159–167.
- PARDUE, M. L., and P. G. DEBARYSHE, 2003 Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu. Rev. Genet.* **37**: 485–511.
- PASYUKOVA, E. G., and S. V. NUZHIDIN, 1993 *Doc* and *copia* instability in an isogenic *Drosophila melanogaster* stock. *Mol. Gen. Genet.* **240**: 302–306.
- PAVLÍČEK, A., J. PACES, D. ELLEDER and J. HEJNAR, 2002 Processed pseudogenes of human endogenous retroviruses generated by LINEs: their integration, stability, and distribution. *Genome Res.* **12**: 391–399.
- PERONNET, F., J. L. BECKER, J. BECKER, L. D'AURIOL, F. GALIBERT *et al.*, 1986 1731, a new retrotransposon with hormone modulated expression. *Nucleic Acids Res.* **14**: 9017–9033.
- PICARD, G., J. C. BRÉGLIANO, A. BUCHETON, J. M. LAVIGE, A. PÉLISSON *et al.*, 1978 Non-Mendelian female sterility and hybrid dysgenesis in *Drosophila melanogaster*. *Genet. Res.* **32**: 275–287.
- RAMOS, K. S., Q. HE, T. KALBFLEISCH, D. R. MONTROYA-DURANGO, I. TENENG *et al.*, 2007 Computational and biological inference of gene regulatory networks of the LINE-1 retrotransposon. *Genomics* **90**: 176–185.
- REBOLLO, R., B. HORARD, B. HUBERT and C. VIEIRA, 2010 Jumping genes and epigenetics: towards new species. *Gene* **454**: 1–7.
- RUBIN, G. M., M. G. KIDWELL and P. M. BINGHAM, 1982 The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell* **29**: 987–994.
- SASAKI, T., H. NISHIHARA, M. HIRAKAWA, K. FUJIMURA, M. TANAKA *et al.*, 2008 Possible involvement of SINEs in mammalian-specific brain formation. *Proc. Natl. Acad. Sci. USA* **105**: 4220–4225.
- SERAFINO, A., E. BALESTRIERI, P. PIERIMARCHI, C. MATTEUCCI, G. MORONI *et al.*, 2009 The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation. *Exp. Cell Res.* **315**: 849–862.
- SHALGI, R., Y. PILPEL and M. OREN, 2010 Repression of transposable elements: A microRNA anti-cancer defense mechanism? *Trends Genet.* **6**: 253–259.
- SHAPIRO, J. A., 1969 Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. *J. Mol. Biol.* **40**: 93–105.
- SHAPIRO, J. A., 2010 Mobile DNA and evolution in the 21st century. *Mob. DNA* **1**: 4.
- SHRIMPTON, A. E., T. F. MACKAY and A. J. BROWN, 1990 Transposable element induced response to artificial selection in *Drosophila melanogaster*: molecular analysis of selection lines. *Genetics* **125**: 803–811.
- SINZELLE, L., Z. IZSVÁK and Z. IVICS, 2009 Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell. Mol. Life Sci.* **66**: 1073–1093.
- SKIPPER, M., U. WEISS and N. GRAY, 2010 Plasticity. *Nature* **465**: 703.
- SOBIESKI, D. A., and F. C. EDEN, 1981 Clustering and methylation of repeated DNA: persistence in avian development and evolution. *Nucleic Acids Res.* **9**: 6001–6015.
- SONG, M., and S. BOISSINOT, 2006 Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene* **390**: 206–213.
- SONG, S. U., T. GERASIMOVA, M. KURKULOS, J. D. BOEKE and V. G. CORCES, 1994 An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev.* **8**: 2046–2057.
- STROBEL, E., P. DUNSMUIR and G. M. RUBIN, 1979 Polymorphism in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in *Drosophila*. *Cell* **17**: 429–439.
- SVANEN, M., 1984 The evolutionary implications of mobile genetic elements. *Annu. Rev. Genet.* **18**: 271–293.
- TIAN, Z., C. RIZZON, J. DU, L. ZHU, J. L. BENNETZEN *et al.*, 2009 Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Res.* **19**: 2221–2230.
- TINYAKOV, G. G., 1939 A highly mutable line from the wild population of *Drosophila melanogaster*. *Dokl. Akad. Nauk SSSR* **22**: 615.
- VARMUS, H., 1988 Retroviruses. *Science* **240**: 1427–1435.
- VENNER, S., C. FESCHOTTE and C. BIÉMONT, 2009 Dynamics of transposable elements: toward a community ecology of the genome. *Trends Genet.* **25**: 317–323.
- VENTURINI, G., E. CAPANNA and B. FONTANA, 1987 Size and structure of the bird genome: II. Repetitive DNA and sequence organization. *Comp. Biochem. Physiol.* **87**: 975–979.
- VIEIRA, C., D. LEPETIT, S. DUMONT and C. BIÉMONT, 1999 Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol. Biol. Evol.* **16**: 1251–1255.

- VOLFF, J. N., 2006 Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* **28**: 913–922.
- VOYTAS, D. F., and F. M. AUSUBEL, 1988 A copia-like transposable element family in *Arabidopsis thaliana*. *Nature* **336**: 242–244.
- WICKER, T., F. SABOT, A. HUA-VAN, J. L. BENNETZEN, P. CAPY *et al.*, 2007 A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**: 973–982.
- WILKINS, A. S., 2010 The enemy within: an epigenetic role of retrotransposons in cancer initiation. *Bioessays* **32**: 856–865.
- WRIGHT, D. A., and D. F. VOYTAS, 1998 Potential retroviruses in plants: Tat1 is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. *Genetics* **149**: 703–715.
- XIONG, Y., and T. H. EICKBUSH, 1988 Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Mol. Biol. Evol.* **5**: 675–690.
- YAMAGUCHI, O., T. YAMAZAKI, K. SAIGO, T. MUKAI and A. ROBERTSON, 1987 Distribution of three transposable elements, *P*, *297*, and *copia*, in natural populations of *Drosophila melanogaster*. *Jap. J. Genet.* **62**: 205–216.
- ZHANG, J., C. YU, V. PULLETIKURTI, J. LAMB, T. DANILOVA *et al.*, 2009 Alternative Ac/Ds transposition induces major chromosomal rearrangements in maize. *Genes Dev.* **23**: 755–765.
- ZIARCZYK, P., and M. BEST-BELPOMME, 1991 A short 5' region of the long terminal repeat is required for regulation by hormone and heat shock of *Drosophila* retrotransposon 1731. *Nucleic Acids Res.* **19**: 5689–5693.