Curr. Issues Mol. Biol. (2002) 4: 65-76.

Transposable Elements and the Evolution of Eukaryotic Complexity

Nathan J. Bowen[1] and I. King Jordan[2]*

[1]Section on Eukaryotic Transposable Elements, Laboratory of Eukaryotic Gene Regulation, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA
[2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

Abstract

Eukaryotic transposable elements are ubiquitous and widespread mobile genetic entities. These elements often make up a substantial fraction of the host genomes in which they reside. For example, approximately 1/2 of the human genome was recently shown to consist of transposable element sequences. There is a growing body of evidence that demonstrates that transposable elements have been major players in genome evolution. A sample of this evidence is reviewed here with an emphasis on the role that transposable elements may have played in driving the evolution of eukaryotic complexity. A number of specific scenarios are presented that implicate transposable elements in the evolution of the complex molecular and cellular machinery that are characteristic of the eukaryotic domain of life.

Introduction

With the recent publication of the human genome draft sequence (Lander *et al.*, 2001; Venter *et al.*, 2001), came the startling revelation that almost half of our genome is composed of the remnants of mobile and semi-autonomous genetic entities referred to collectively as transposable elements. For example, members of a single family of transposable elements, known as LINEs, were shown to make up fully 21% of the human genome sequence. By comparison, protein-coding sequences make up only ~1% of the genome. Furthermore, the total estimate of 45%, for the fraction of the genome made up by transposable elements, is certainly a significant underestimate as many of these rapidly evolving elements are likely to have changed beyond recognition. And so it is that these previously marginalized and lightly regarded elements have been thrust into the spotlight of biological science. In fact, the abundance and ubiquity of transposable elements had already been very much established and was well appreciated among the community of biologists dedicated to their study. The implication of this understanding, now widely shared, is simply that one cannot begin to fully comprehend the evolution, the organization, the structure and even the function of eukaryotic genomes without a real understanding of transposable elements.

This review is concerned mainly with the evolutionary implications of transposable elements. An accumulation of molecular evidence over the last decade or so has unequivocally demonstrated a myriad of ways that transposable elements can affect the evolution of the organisms in which they reside (their host species). Important as they are, these findings will be covered in a mostly cursory manner since they have been reviewed in detail elsewhere (McDonald, 1993; Wessler *et al.*, 1995; Britten, 1996; Kidwell and Lisch, 2001; Nekrutenko and Li, 2001). A greater emphasis will be placed on more speculative notions concerning the ways in which transposable elements may have influenced the evolution of complexity in eukaryotes. Often in these cases, the connections between the molecular data bearing on transposable elements and the evolutionary inferences drawn from them are preliminary and as such should be treated judiciously. We find that it is precisely the ambiguity that comes with this, and the associated latitude for differing interpretations, that renders these perspectives compelling and noteworthy. However, the inclusion of any particular transposable element related evolutionary scenario here is not to be taken as a statement of its validity or lack thereof for that matter. Rather, our hope is that a presentation of these more formative ideas may play some role in stimulating new ways of thinking about old questions, and perhaps more importantly, aid in the development of novel testable hypotheses.

Transposable Elements Defined

Transposable elements are repetitive genomic sequences that are able to move (transpose) from one chromosomal location to another. In so doing they often replicate themselves. The replicative capacity of transposable elements is essence of their evolutionary success and has important implications for the dynamics between elements and their hosts. Transposable elements are classified based upon their mechanism of transposition (Finnegan, 1992) as well as by comparison of their genomic structures and sequences (Figure 1). Class I elements, or retroelements, transpose via the reverse transcription of an RNA intermediate, and class II elelments, or DNA elements, transpose directly as DNA using a cut and paste mechanism.

Class I elements can be divided into several subclasses commonly referred to as SINEs, LINE-like elements and long terminal repeat (LTR) retrotransposons

*For correspondence. Email jordan@ncbi.nlm.nih.gov; Tel. 301-594-5714; Fax. 301-480-9241.

© 2002 Caister Academic Press

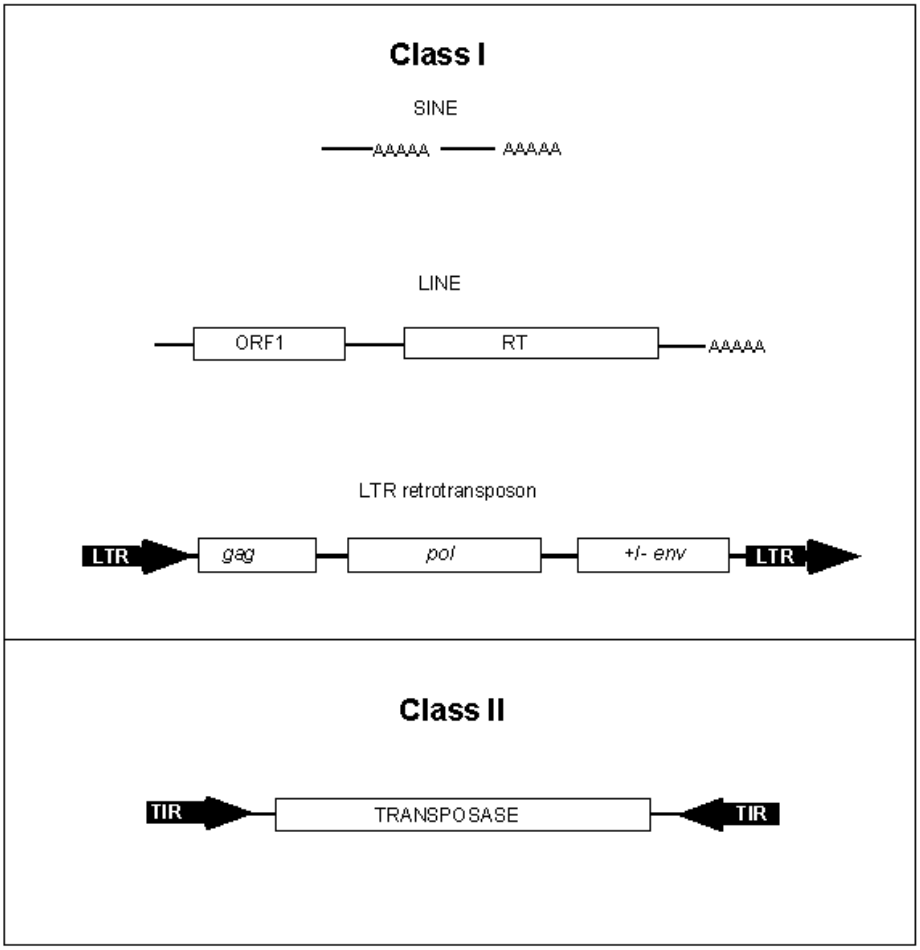


Figure 1. Schematic of the genomic structure of class I (SINE, LINE-like and LTR retrotransposon) and class II transposable elements. Abbreviations: RT - reverse transcriptase, LTR - long terminal repeat, TIR - terminal inverted repeat.

(Figure 1). The SINEs or retroposons are unique among these groups in that none of them encode their own transposition machinery and are thus retrotransposed in trans by enzymatic machinery encoded elsewhere. The Alu elements, which make up more than 10% of the human genome, are the most well known member of this subclass (Mighell *et al.*, 1997). Non-LTR elements, or LINE-like elements as they are colloquially known, do encode the enzymes used in their transposition. They usually possess two open reading frames and an internal promoter. Reverse transcription of these elements is often prematurely terminated resulting in the formation of many truncated non-autonomous copies. Both SINEs and LINEs contain A rich regions in their 3' ends that are used for priming the reverse transcription reaction. LTR retrotransposons have a genomic structure that is virtually identical to that of retroviruses, and in fact they are close evolutionary relatives of retroviruses. LTR elements are flanked by two identical long terminal repeats and usually have 1 - 3 open reading frames that encoded structural and enzymatic proteins involved in retrotransposition (Figure 1). Reverse transcription of LTR retrotransposons takes place in a viral-like particle. Most LTR retrotransposons are autonomous but there are a few cases of non-autonomous LTR elements that do not encode their own transposition machinery.

Class II elements are related to bacterial transposons and have terminal inverted repeats that flank an open reading frame encoding a transposase enzyme (Figure 1). The transposase enzyme binds class II element DNA sequences at or near the inverted repeats and catalyzes transposition by cutting out the element sequence and inserting it at a new location.

Evolutionary Dynamics of Transposable Elements

In a remarkable piece of genetic detective work, Barbara McClintock first inferred the presence of mobile genetic elements in maize in the late 1940s (McClintock, 1948). Biologists would have to wait almost 20 years for the original molecular characterization of a transposable element, a class II bacterial insertion element (Jordan *et al.*, 1967). After these initial discoveries of transposable elements, there was much speculation about their evolutionary origin.

Most of the discussion was centered on the possible roles that elements played for their host genomes. This speculation was in line with the 'phenotypic paradigm' of the neo-Darwinian theory that holds that genes ensure their survival and representation in subsequent generations by providing a selective advantage for their host organism (Doolittle and Sapienza, 1980). Following this idea, it was reasoned that the presence of transposable elements must be due to some function they perform for their host genomes. Two seminal papers published simultaneously in Nature in 1980 directly challenged this line of thinking (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). In what amounted to a true paradigm shift, the authors of these papers concluded that the emergence and spread of transposable elements could be explained solely by their ability to replicate themselves in the genome. The replicative capacity of transposable elements provides them with a biased transmission relative to host genes. Because they can out-replicate their hosts, the evolutionary success of these elements is largely irrelevant to any selective advantage provided to their host genome. It was even demonstrated theoretically that transposable elements could spread and persist in natural populations in the face of a selective disadvantage for their host organisms (Hickey, 1982). These findings form the core of the selfish DNA theory of transposable elements. This theory emphasizes the parasitic nature of transposable elements. The underlying logic and coherence of this theory are so compelling as to be undeniable. However, the acceptance of the theory led to a drastic stance on the evolutionary significance of transposable elements. It was concluded that transposable elements are merely genomic parasites and their selfish nature precludes them from playing an important role in genome evolution. Consider this quote (Orgel and Crick, 1980):

> When a given DNA, or class of DNA, of unproven phenotypic function can be shown to have encoded a strategy (such as transposition) which ensures its genomic survival, then no other explanation for its existence is necessary. The search for other explanations may prove if not intellectually sterile, ultimately futile.

Taking the selfish DNA theory of transposable elements to such an extreme conclusion had the effect of discouraging evolutionary studies into the ways that elements could influence their host genomes. Despite this chilling effect, an accumulation of molecular evidence would soon reveal that, while transposable elements are by and large genomic parasites, they have been co-opted many times and in a number of different ways to serve the interests of their hosts.

Transposable Element Influences on Host Evolution

Evolution is a tinkerer (Jacob, 1977). As opposed to creating new forms *de novo*, it will more often re-arrange whatever materials are at hand to create novelty. Transposable elements, abundant and ubiquitous as they are, can serve as ideal genetic building blocks with which evolution can tinker to create modified forms. Molecular investigations into many different systems have led to example after example of ways that transposable elements have been so used and thus influenced the evolution of their host organisms. Wolfgang Miller famously dubbed this phenomenon 'molecular domestication' (Miller *et al.*, 1997). Numerous cases of molecular domestication have been reviewed at length elsewhere (McDonald, 1993; Wessler *et al.*, 1995; Britten, 1996; Kidwell and Lisch, 2001; Nekrutenko and Li, 2001), and below we will merely attempt to familiarize the reader with several broad classes of effects that transposable elements have been demonstrated to exert on their host genomes. These include the influences of transposable elements on gene regulatory evolution and the role of elements in protein-coding sequence evolution.

In order to facilitate their transcription, transposable elements often encode their own promoter sequences. For example, the long terminal repeats of LTR retrotransposons contain promoter and enhancer sequences in addition to transcription initiation and termination sites. LINE-like elements often have internal pol II promoter sequences. If a transposable element inserts in a gene-containing region, the potential exists for its promoter and enhancer sequences to influence the regulation of the nearby host gene (White *et al.*, 1994). In fact, a number of cases exist where it can be shown that: 1 - an element inserted in a gene region, 2 - it has been constrained by natural selection and 3 - it affects the transcriptional regulation of the host gene (Britten, 1996). For example, the expression of the mouse sex-linked protein encoding gene is influenced by the insertion of an LTR retroelement in its 5' region (Stavenhagen and Robins, 1988). Regulatory sequences in the LTR have been conserved by selection and confer a distinct androgen dependent male specific expression pattern on this gene. Whats more, the sex-linked protein has been shown to play an important functional role in the mouse (van den Berg *et al.*, 1992). This is a classic example of how transposable element sequences can influence the regulatory evolution of host genes. The abundance of transposable element sequences, taken together with their mobility, signifies the existence of a vast reservoir of mobile promoter sequences. The ability and number of different ways these promoter sequences can potentially affect host gene regulation is seemingly unlimited.

The idea of the regulatory potential of repetitive element sequences has actually been around for some time (Britten and Davidson, 1971). What has been less understood is the effect of transposable elements on the evolution of protein-coding regions. Several recent publications indicate the extent to which transposable elements insertions within protein-coding regions have resulted in the creation of novel genes (Brosius, 1999; Makalowski, 2000; Li *et al.*, 2001). For example, an anlaysis of ~14,000 human genes revealed that 4% contained transposable element insertions (Nekrutenko and Li, 2001). The majority of these insertions were into pre-existing introns and many have resulted in the creation novel exonic sequences. This same study also found more than 100 cases where transposable element insertions are responsible for the divergence between orthologous genes from different species. Thus transposable element

insertions may not only represent a mechanism for the formation of new genes but may also play a prominent role in species diversification.

One unique case of molecular domestication that does not fit neatly into either of the above categories, gene regulatory or coding evolution, is worth noting. This is the case of the Het-A and TART elements in *Drosophila melanogaster*. These are LINE-like elements that have been co-opted by the host genome to maintain the integrity of the telomeres (Levis *et al.*, 1993; Sheen and Levis, 1994; Pardue and DeBaryshe, 1999). *D. melanogaster* does not encode telomerase and so can not use this enzyme to catalyze regeneration of chromosome ends that are progressively shortened during rounds of DNA replication. Amazingly, the *Drosophila* genome has domesticated two transposable elements, Het-A and TART, to perform this function. These elements insert specifically at chromosome ends. By repeatedly transposing and inserting at the telomeres, they are able to extend the chromosome ends and offset the loss of DNA associated with replication. This is a remarkable example of host-element co-evolution resulting in mutualism. The host has enlisted the elements to solve a problem critical to its survival. The elements, on the other hand, ensure their survival by continuing to replicate while simultaneously endearing themselves to their host genome.

The Complexity Trend

Now that we have reviewed a sample of the current established thinking on the evolutionary influences of transposable elements, we will move to describe some more speculative and/or provocative examples of how elements may have influenced the evolution of eukaryotic complexity. In a general sense, organismic complexity can be defined with respect to the number of different physical parts, and by extension the number of interactions between different parts, of an organism. For example, organismic complexity is very often measured by the number of different cell types (Carroll, 2001). Cell type number, like other measures of organismic complexity, has obviously increased over evolutionary time. However, it is necessary to further explore the question of the evolution of complexity to specifically address whether there is in fact a progressive trend toward greater complexity in evolution. The question of a trend towards greater complexity is a sensitive one that vexes many evolutionary biologists. This may be because there are some decidedly non-scientific implications associated with the notion of a complexity trend. For example, before Darwin a popular conception of the organization of life was the so-called 'Great Chain of Being' (Futuyma, 1986). According to this scheme, nature was organized in a strict hierarchy with lower (less complex) organisms near the bottom giving way to progressively higher (more complex) organisms near the top with humans at the pinnacle. This hierarchy, which reflected God's plan, was considered to be fixed and unchanging; evolution was unthinkable. The role of biology was to discern the identity and order of the links in this chain. This task was undertaken explicitly to reveal the wisdom of the Creator. Thus Linnaeus presented his famous classification system *ad majorem Dei gloriam*, 'for the greater glory of God.' Implicit in this decidedly non-evolutionary scheme was the concept of an upwards teleological thrust that manifested itself as a trend towards ever increasing complexity. In short, the postulation of an inherent trend towards increasing biological complexity is often thought to imply the existence of ultimate causes.

Despite the teleological connotations of a progressive complexity trend, it is difficult to ignore the unmistakable observation that life has increased in complexity since its origins (Carroll, 2001). Since lifes' relatively humble beginnings, evolution has produced more and more complex organisms. Not only is this the case, but the average complexity of life appears to have increased as well. The question is whether this increase in complexity reflects a truly active and directional trend in evolution. Stephen Jay Gould addressed just this question in his book Full House (Gould, 1996). Gould contends that the evolutionary increase in organismic complexity is actually passive and random rather than active and directional. The apparent increased complexity merely reflects a passive tendency to evolve away from the initial lower bound of organismic complexity. In other words, the appearance of an increase in complexity actually results from an increase in the variance of complexity. The direction towards greater complexity is imposed by the initial boundary of low complexity ancestral forms (i.e. bacteria). In support of this notion Gould points out that while the mean of organismic complexity has increased the mode has not. Judged by a number of different criteria (e.g. number, biomass, niche diversity), bacteria remain the dominant life form on earth. Since extant bacteria are presumably no more complex than ancestral bacteria, the mode of organismic complexity has not increased. An active directional or progressive trend towards increased complexity should result in the replacement of primitive low complexity forms with derived higher complexity forms, resulting in a change in the mode of complexity. Since this has clearly not occurred, the global trend towards increasing organismic complexity is considered to be passive or random. To Gould, the appearance of an increasing complexity trend simply reflects a myopic focus on extreme values. It is worth noting, however, that while the global trend towards increased complexity may indeed be passive, there is evidence for active complexity increases within related groups (clades) of organisms (Carroll, 2001). The relationship between transposable elements and eukaryotic complexity will be considered in light of this controversy. We will provide evidence that suggests ways in which transposable elements may have played an active role in driving the increased organismic complexity characteristic of the eukaryotic domain of life.

Organismic Complexity

Let us first consider the relationship between organismic complexity and transposable elements in the general sense. Robert Wright recently treated the subject of increasing complexity at length in his book Non-Zero: The

Logic of Human Destiny (Wright, 2000). The term 'non-zero' is taken from game theory and refers to 'non-zero sum' games where players' interests overlap, as opposed to 'zero-sum' games where the players' interests are directly at odds to one another. Wrights' thesis is that a combination of zero-sum and non-zero sum interactions inevitably leads to greater complexity, whether in human societies or the natural world. Wright gives a number of examples where organisms interact in ways that are ultimately, if not proximally, non-zero sum and describes how these interactions may lead to increases in complexity. For example, predator and prey species have been shown to engage in evolutionary 'arms races' where the prey evolves ways to avoid capture by the predator and the predator evolves ways to circumvent these defensive measures. Wright sees this as the source of a directional or active trend toward increasing organismic complexity and takes aim at Gould's contention that the global increase in organismic complexity reflects a random or passive trend. Wright points out that Gould emphasizes that natural selection can only adapt organisms to their local environments. So it is the effectively random changes in local environment that organisms track via natural selection. Therefore, the trajectory of the changes wrought by natural selection should be random and should not lead to a directional trend of increasing complexity. However, Wright notes that organisms' environments consist of not only the physical elements but also, and perhaps more importantly, of other organisms. Since the average complexity of the organismic component of the environment is increasing due to species' interactions, environmental tracking by natural selection may in fact lead to a progressive increase in complexity.

Transposable elements fit nicely with Wright's more inclusive concept of organismic environment by extending this notion inward. With respect to element-host interactions, the environment that natural selection must concern itself with consists of not only the physical environment and other species but also the micro-environment of the genome itself. As described above, transposable elements have a semi-autonomous existence within the genome in the sense that they are largely responsible for their own replication and may not have the same evolutionary interests as their host genome. Indeed, transposition is often harmful to the host and natural selection will favor host organisms that can devise ways to mitigate the deleterious consequences of element insertion (Charlesworth *et al*., 1994). This is manifest in various repression mechanisms that effectively reduce transposition rates (Miller *et al*., 1997). Meanwhile, transposable elements are under selective pressure to escape these repression mechanisms (Bestor, 1999). One radical solution that some elements have evolved is to laterally transfer across species boundaries (Lohe *et al*., 1995; Kidwell and Lisch, 1997). Introduction into a novel genomic environment may allow for an enhanced transposition rate since a newly infected host has not yet been pressured to evolve repression mechanisms. But transposable elements, like other parasites, are faced with an even more complicated evolutionary scenario than the one outlined above. To guarantee their representation in subsequent generations, they should be able to transpose and thus replicate within the genome. However, they cannot transpose so effectively as to destroy the host species in which they reside. Elements have evolved a number of mechanisms that allow them to strike this delicate selective balance. For example, P elements in *Drosophila* use alternative splicing to encode a transposase enzyme in the germ-line and a repressor enzyme in somatic cells (Rio, 1990). In this way transposition is limited to germ-line cells. This eliminates the deleterious effects of somatic transposition, while maximizing the chances that newly transposed elements will be represented in subsequent generations. Other elements, like LTR retrotransposons in plants and fungi, appear to have evolved a preference for integrating in intergenic regions (SanMiguel *et al*., 1996; Kim *et al*., 1998; Behrens *et al*., 2000). This strategy can also serve to blunt the harmful effects of transposition. Obviously, the co-evolutionary dynamic between host and elements is an intricate one. Elements can clearly be considered to play both zero sum and non-zero sum games with their host organisms. Such a back and forth between elements and their hosts may very well lead to a spiral of increasing complexity much as Wright envisions for interacting organisms. Could elements be one source of an evolutionary 'creative thrust' (Wright, 2000), in a sense both directed and active, towards ever increasingly complex life forms? It may very well be the case. Several examples where transposable elements might have facilitated the evolution of the molecular machinery characteristic of and essential to eukaryotic complexity are presented below.

The Utility of 'Junk' DNA

Genome size as measured by the number of base pairs (bp) shows an astonishing 80,000-fold variation among eukaryotes (Li and Graur, 1991). Interestingly, variation in eukaryotic genome size does not correlate with either genomic (number of genes) or organismic (number of cell-types) complexity. For instance, genome size varies the most among single celled protozoans and a number of these organisms have larger genomes than any mammal. The absence of a correlation between genome size and the amount of genetic information encoded in a genome is known as the C-value paradox. This paradox is partially explained by the fact that, for most eukaryotic genomes, the vast majority of sequence does not correspond to any gene or encode any protein. In fact, eukaryotic genomes consist mainly of vast amounts of non-coding intergenic sequences know as 'junk' DNA. It turns out that most of this sequence is derived from transposable elements. For example, in humans ~45% of genomic sequence has recognizable similarity to known transposable element sequences (Lander *et al*., 2001). Most of the remaining non-coding human DNA is also presumably derived from transposable elements but has diverged beyond recognition. Even more remarkable is the finding that species of the plant genus *Lilium* have huge genomes (36 x 10^9 bp) that are made up of ~99% transposable element sequences (Leeton and Smyth, 1993).

The moniker junk DNA refers to the fact that there is

no known or established function for this abundance of sequence. This class of genomic DNA is widely assumed to be basically worthless and even dispensable. However, the fact that most of the so-called junk DNA is made up of transposable elements provides a compelling explanation for its presence. It can be considered that this excess DNA exists because of the selfish drive of elements to replicate themselves in the genome and the inability of the host to completely rid itself of such sequence. The completion of a recent eukaryotic genome sequencing project provided some intriguing results that challenge this perspective on the role of non-coding intergenic DNA.

Cryptomonads are chromphyte algae that retain an additional miniature nucleus know as a nucleomorph. This nucleomorph is the remnant nucleus of an endosymbiont red algae. All three nucleomorph chromosomes from the cryptomonad *Guillardia theta* were recently sequenced (Douglas *et al*., 2001). The diminutive 551-kb nucleomorph genome is the most gene-dense eukaryotic genome known. There exists a paucity of non-coding regions including intergenic spacer regions and spliceosomal introns in this genome. In addition, there are no known transposable element sequences found within the nucleomorph genome. However, the larger co-existing genome of *Guillardia theta* does contain abundant non-coding DNA and transposable element sequences.

The effective elimination of transposable element sequences by the nucleomorph genome was taken to indicate that selection against non-functional DNA is effective enough to completely eliminate it (Cavalier-Smith and Beaton, 1999). The converse of this argument is that the persistence of substantial amounts of non-coding DNA and transposable element sequences in the larger co-existing genome, as well as virtually all other eukarytoic genomes, suggests that selection based on some functional constraint must play a role in maintaining it. If selection is indeed conserving the presence of junk DNA then there must be some function for it. The authors of this idea state their case concisely as follows (Douglas *et al*., 2001):

> The marked contrast between the effective elimination of non-coding DNA from cryptomonad nucleomorphs and the accumulation of vast amounts of non-coding DNA in coexisting cryptomonad nuclei indicates that nuclear non-coding DNA in general is functional and positively selected, and is not purely selfish or junk.

What could the function of this DNA be? The skeletal DNA hypothesis is an attempt to postulate a cellular role for so-called junk DNA (Cavalier-Smith and Beaton, 1999). According to this hypothesis, optimal cellular function requires a relatively constant cytonuclear ratio. In other words, cell volume and nuclear volume (genome size) must be positively correlated. Despite the lack of correlation between genome size and complexity, genome size and nuclear volume do increase with increasing cellular volume. In this sense, the increase in genome size mediated by an accumulation of non-coding intergenic DNA, made up mainly of transposable elements, may be due to a skeletal role that this DNA plays in maintaining critical nuclear volume. This is one example of a global functional role that transposable element sequences may play for their host genomes.

Genomic Complexity

Transposable elements may have played an important role in facilitating two major evolutionary transitions that are marked by considerable increases in genome complexity. Genomic complexity is often approximated in a straightforward way as the number of genes in a genome. Relying on preliminary estimates of gene number from the pre-genomics era, Adrian Bird pointed out that both the origin of eukaryotes and the origin of vertebrates were marked by sudden increases in gene number (Bird, 1995). These increases are likely to have provided the additional coding capacity that resulted in the emergence of many of the novel structures, functions and pathways characteristic of the newly evolving lineages. However, these increases in coding capacity also necessitated the evolution of global regulatory mechanisms. Without efficient global gene regulation, gene silencing mechanisms in particular, spurious transcription from numerous poorly regulated genes would cause an untenable amount of noise in the genetic programs of the new lineages. Such lineages would be unlikely to survive and thrive without evolving a way to overcome the regulatory challenges posed by major increases in gene number. With this in mind, Bird argued that the evolution of distinct epigenetic silencing mechanisms allowed for the increases in gene number associated with each major macroevolutionary transition: chromatin formation at the prokaryote-eukaryote transition and methylation at the invertebrate-vertebrate transition.

Left unexplained in this scheme however, was the source of the proximate selective forces that could have encouraged the initial formation of global silencing mechanisms. The evolution of such systems would seem to be counterintuitive in light of the fact that global silencing would most likely be disadvantageous in the short-term, from the perspective of more or less well-regulated host genes, and thus selected against. Since natural selection lacks the foresight necessary to endure short-term setbacks in hopes of more substantial long-term gains, the evolution of global silencing would seem to be precluded. Into this breach stepped John McDonald, who marshaled data that implicated transposable elements in driving the evolution of global silencing mechanisms (McDonald, 1998). McDonald pointed to a growing body of evidence that supports the notion that specific global repression mechanisms arose as adaptive responses to the presence of notoriously prolific transposable element sequences (Dorer and Henikoff, 1994; Martienssen, 1996; Yoder *et al*., 1997; Jensen *et al*., 1999; Matzke *et al*., 1999). As described previously, transposable elements can easily out-replicate the genomes in which they reside and will spread and multiply in the genome often with disastrous consequences for their hosts. Accordingly, host genomes have a compelling selective interest in reducing transposition rates and have evolved a number of different ways of doing so. Chromatin formation and methylation are two global silencing mechanisms that are thought to

have evolved to repress the activity of transposable elements. In this way, the selective pressure exerted on genomes by transposable elements may have played a major role in facilitating the increases in genome complexity that characterized the evolutionary transitions to eukaryotes and vertebrates respectively (McDonald, 1998).

These two associated hypotheses, while very attractive, need to be reconsidered in light of recent genomic data. Birds' hypothesis rested on what he referred to as a 'quantum' increase in gene number at the invertebrate-vertebrate transition. Invertebrate genomes were considered to contain 7,000 - 25,000 genes while vertebrate gene numbers were estimated to range from 50,000 - 100,000 (Bird, 1995). While these figures do imply a sudden increase in gene number at the invertebrate-vertebrate transition, recent genomic data give lie to this supposition. For example, one of the surprises of the human genome project was the revised estimate of ~30,000 human genes (Lander *et al*., 2001; Venter *et al*., 2001). This is substantially lower than most previous estimates, including the ones Bird used, that placed this figure around 100,000. The *C. elegans* genome project revealed a gene number ~20,000 (Consortium, 1998). These two estimates are surprisingly close given the obvious differences in phenotypic complexity between the two organisms as well as the amount of time elapsed since their divergence. Furthermore, the change in gene number associated with the emergence of vertebrates was quite probably not as sudden or dramatic as previously thought. However, the emergence of a novel global silencing mechanism driven by transposable elements may still have been critical in facilitating an increase in the complexity of the genetic programs of emerging vertebrates. In this case though, the critical parameter may not have been gene number per se.

It may actually be an increase in the number of different proteins that are encoded by the genomes of vertebrates that demands the presence of global regulatory mechanisms. Alternative splicing is known to be particularly prevalent in the human genome (Lander *et al*., 2001). An examination of human chromosome 22 yielded evidence that 59% of the genes had alternative splice variants with an average of 2.6 distinct transcripts per gene. Chromosome 19 showed an average of 3.2 distinct transcripts per gene. This was substantially higher than what was seen in *C. elegans* where 22% of genes were found to have alternative transcripts and the average number of distinct transcripts per gene was found to be 1.3. From this it has been inferred that the level of alternative splicing is substantially higher for vertebrates than for less complex eukaryotes. If this proves to be the case, then there may indeed have been a 'quantum' increase in the number of gene products at the vertebrate transition if not in actual gene number. Such an increase in alternative splicing and resulting gene product number would also seem to demand an increase the efficiency of global gene silencing provided by methylation. However, it is not currently clear whether the apparent increase in alternative splicing is merely an artifact of the tremendous effort put into EST characterization for vertebrate genomes and/or the small number of genomes systematically characterized with respect to this phenomenon.

Table 1. Gene density for eukaryotic genomes.

Species	Number of base pairs per gene[a]
Guillardia theta nucleomorph	977
Encephalitozoon cuniculi	1,025
Saccharomyces cerevisiae	1,923
Schizosaccharomyces pombe	2,481
Caenorhabditis elegans	6,180
Drosophila melanogaster	9,000
Homo sapiens	110,000

[a] Calculated by dividing genome size (base pairs) by the number of genes

Nuclear Architecture and Gene Regulation

Classic models of gene regulation invoke the interaction of cis nucleotide sequence elements and trans acting protein factors. A more modern molecular view has added a global perspective to the understanding of gene regulation. For example, the importance of epigenetic phenomena such as methylation and chromatin modification is now widely recognized (Wolffe and Matzke, 1999). In addition, the influence of nuclear architecture on transcriptional regulation has attracted much attention as of late. A consensus is emerging that describes the nucleus as containing discrete regions with specific and uniform transcriptional and replicational properties (Berezney and Wei, 1998). There are a number of lines of evidence that implicate transposable elements in the evolution of such a nuclear architecture based regulatory system.

The partitioning of chromosomal segments into different nuclear regions is governed by the attachment of specific genomic sequences, known as matrix attachment regions, to the nuclear matrix. Matrix attachment regions are thought to correspond to origins of replication (ORC) (Berezney and Wei, 1998). Interestingly, recent work in *S. cerevisiae* shows that the locations of yeast replication origins are astonishingly coincident with the location of Ty element LTRs (Wyrick *et al*., 2001). Another connection between transposable elements and matrix attachment regions is manifest by the *Drosophila* retroelement gypsy. Gypsy contains a matrix attachment region that serves to insulate enhancers from promoters by sequestering them in separate nuclear regions (Nabirochkin *et al*., 1998). It does this by aggregating near the nuclear periphery with other insulators (Gerasimova *et al*., 2000).

A general effect that transposable elements have that is related to nuclear architecture is the partitioning of the genome into discrete sections. This could facilitate a more specific level of nuclear architecture related gene regulation. One common aspect of the genomes of most complex eukaryotes is the presence of relatively large intergenic regions. This is reflected in the values of the gene density parameter for completely sequenced eukaryotic genomes (Table 1). The gene density parameter is calculated by dividing the size (# bp) of the genome by the total number of genes. For example, given their obvious differences in organismic complexity, *C. elegans* and *H. sapiens* have surprisingly close gene numbers. However, the gene density of the *H. sapiens* genome is much lower than that of *C. elegans*. In this sense, gene density may

be considered a more accurate measure of genomic complexity. Consistent with this idea, *D. melanogaster* also has a lower gene density than *C. elegans* despite the counter-intuitive finding that less organismically complex *C. elegans* has more genes than *D. melanogaster*. Lower gene density may reflect an increase in organismic complexity caused by more elaborate regulatory networks. Such regulatory networks could well be mediated by the presence of nuclear matrix attachment regions that are known to be found in both intergenic regions and transposable element sequences. Binding of these regions to the nuclear matrix forms chromosomal loops that are thought to be independently regulated. This regulation is achieved by the occupation of specific nuclear regions that correspond to zones of transcription (Berezney and Wei, 1998). The genomes of eukaryotes with low gene density and long intergenic regions will contain loops with a few or even single genes. This gene specific nuclear localization can provide an additional level of regulatory specificity on top of the classical cis-element trans-factor model of gene regulation. The increased specificity and flexibility of nuclear localization based regulation in low density genomes may entail one mechanism that leads to increased organismic complexity. Furthermore, the presence of transposable elements in intergenic regions as well as the connections between transposable elements and matrix attachment regions points to an important role for these elements in contributing to the evolution of this type of gene regulation.

Sexual Reproduction

Sexual reproduction is a leit motif of eukaryotic biology. Recent work suggests that the molecular machinery that drives reductive cell division (meiosis) emerged very early in eukaryotic evolution (J. Logsdon, pers. comm.). The existence of sexual reproduction represents a paradox for evolutionary biologists as sex exacts a substantial evolutionary cost. This is because an organism that reproduces asexually (clonally) creates exact genetic copies of itself and thus leaves twice as much genetic material per offspring as a sexually reproducing organism. This would seem to be an almost insurmountable fitness advantage for a newly evolving sexually reproducing lineage to overcome. Evolutionary biologists have long held that in order for sexual reproduction to evolve in the face of this cost it must have provided some tangible and considerable benefits. An example of a long term benefit that may have allowed sex to persist in eukaryotes is the ability of sexually reproducing populations to recombine beneficial mutations leading to an increased rate of evolution (Fisher, 1930; Muller, 1932). However, such a long-term advantage would not be sufficient to allow for the emergence of sex in a population. Any model that explains the evolution of sex must account for the short-term advantage of sex (Maynard Smith, 1978). More recent work has demonstrated that sex and recombination provide a short-term advantage in the elimination of deleterious variants from the population (Kondrashov, 1988; Barton and Charlesworth, 1998). Although these beneficial effects are more than sufficient to explain the evolution of sex despite its cost, evolutionary biologists have yet to agree on a general model that explains the evolution and persistence of sexual reproduction among eukaryotes.

There also exists an interesting relationship between the existence of sexual reproduction and the spread of transposable elements. Hickey demonstrated that sexual reproduction greatly enhances the probability that a transposable element will become fixed in a population (Hickey, 1982). This is due to the fact that transposable elements can increase at twice the Mendelian rate of host genes when uninfected genomes (containing no elements) are always infected when crossed to infected genomes. Given at least one transposition event per generation, elements will become fixed as long as the coefficient of the selection cost imposed by the element is less than 1/2. So under conditions of sexual reproduction, there is a two-fold fitness cost to begin with and there can be another almost two-fold penalty associated with transposition. His awareness of this double penalty for sex, as well as the benefit sex provides to transposable elements, led Hickey to suggest that perhaps transposable elements drove their hosts to evolve sexual reproduction to aid in the elements spread.

It is difficult to formally substantiate such a proposal. However, it is interesting to consider how this scenario relates to what is known about the evolutionary dynamic between host and elements. Transposable elements engage in a complex and fluid evolutionary relationship with their hosts. While elements have clearly been shown to perform a number of functions for their hosts, their presence can be explained strictly in terms of their parasitic behavior. As genomic parasites, elements must walk a fine line with their hosts. The elements need to out replicate the hosts in order to ensure their survival. Meanwhile, the hosts' interests are served by repressing the elements and thus minimizing the deleterious effects of transposition. However, if elements evolve to transpose too efficiently they will become so virulent as to eliminate the host and destroy their own evolutionary vehicle. Earlier we discussed some mechanisms that elements have evolved to strike this evolutionary balance. Sexual reproduction represents a single mechanism by which both host and element can achieve and benefit from such a balance. As Hickey proposed, the evolution of sex would have served to virtually guarantee the evolutionary success of elements. At the same time, sexual reproduction is very effective at helping to eliminate just the type of deleterious mutations that elements often cause. This aspect of sex clearly aids the host but can also be considered to aid the element as it keeps the host viable, allowing for the persistence of its elements. That sexual reproduction can simultaneously benefit both host and element in this way is intriguing and does seem to suggest, at the least, the existence of an intimate relationship between the evolution of sex and the presence of transposable elements. Sexual reproduction is widely recognized as contributing to increased organismic complexity by providing for greatly enhanced levels of variation.

The Vertebrate Immune System

One of the unique and defining characteristics of vertebrate cellular biology is the specific immune system. Vertebrate immune systems have evolved a way to mount specific responses, each precisely tailored to a particular pathogen, via the production of immunoglobulin (Ig) receptors and antibodies. The Ig proteins are expressed in B-lymphocytes and T-cells and encoded by complex loci (V(D)J genes) made up of a series of discrete coding segments (L-V, D, J and C). During lymphocyte differentiation these segments are recombined into combinations containing a mix of segments from different loci, and during a specific immune response the rearranged genes undergo somatic hypermutation. These processes result in an incredibly diverse repertoire of different Ig proteins. This diversity is a critical component of the vertebrate immune system. There is compelling evidence that illuminates the connection between transposable elements and the evolution of the V(D)J recombination process. In addition, a more controversial hypothesis points to a role for transposable elements in the evolution of somatic hypermutation.

Recombination of the V(D)J locus is catalyzed by two enzymes RAG1 and RAG2. The catalytic mechanism by which RAG1 and RAG2 affect recombination was found to be so similar to the transposition mechanism encoded by class II DNA elements (Lewis and Wu, 1997) that it suggested they shared a common evolutionary origin (Spanopoulou *et al*., 1996). Pursuant to this idea, experimental evidence revealed that RAG1 and RAG2 together could in fact catalyze transposition in vitro (Agrawal *et al*., 1998; Hiom *et al*., 1998). Taken together these findings strongly suggest that a transposable element provided the enzymatic machinery pre-requisite for the evolution of V(D)J recombination. This striking example of molecular domestication is entirely consistent with the idea that evolution often co-opts pre-existing forms to use as the building blocks of novel molecular systems.

Edward Steele and colleagues have also proposed a role for transposable elements in the evolution of the Ig component of the specific immune system (Steele *et al*., 1998). Their somatic mutation hypothesis is a provocative one that has yet to receive wide acceptance in the scientific community. The postulated mechanism involves reverse transcription, with enzymatic activity provided by host retroelements, of re-arranged V(D)J pre-mRNAs. The cDNAs that result from this process should be highly mutated since the production of both the mRNA and cDNA are error prone. These cDNAs may then replace the resident V(D)J genes via homologous recombination. This process is envisioned to take place in the sequestered environment of the 'mutatorosome' that contains the reverse transcriptase activity and ensures that only V(D)J genes are so mutated. This specificity is thought to be achieved by specific binding of V(D)J pre-mRNAs, based on their secondary structure, by the protein components of the mutatorosome.

As evidence for this scenario, the authors refer to a number of different sources (Steele *et al*., 1997; Blanden *et al*., 1998a; Blanden *et al*., 1998b; Blanden and Steele, 1998). The basic idea is that analysis of variation among V(D)J genes reveals patterns that favor the hypothesis of somatic mutation followed by reverse transcription and integration by homologous recombination. For example, the high mutation rates are consistent with the action of RNA polymerase and reverse transcriptase. The reverse transcriptase model also explains the boundaries of somatic hypermutation which is limited to a discrete region encompassing the V(D)J genes. In addition, there is evidence that suggests a necessary association between somatic hyptermutation and homologous recombination as proposed by the reverse transcriptase model.

Perhaps the most controversial aspect of this hypothesis is a proposed soma-to-germline feedback mechanism. The mutated and selected genes from B cells are also thought to integrate into germ cell DNA via reverse transcription. The reverse transcriptase activity is thought to be provided by enodgenous retroviruses. The known activity of reverse transcriptase in germ cells and the pattern of variation of germ cell antibody genes are both consistent with this hypothesis. In fact, all of the various molecular mechanisms invoked in their hypothesis are well supported by the scientific literature. However, due to a lack of direct experimental evidence for this hypothesis, the question remains as to whether these mechanisms do in fact interact in a way that is consistent with reverse transcriptase mediated somatic hypermutation and soma-to-germline feedback. If this proves to be the case then it would have profound implications for evolutionary theory. Such a mechanism would imply a Lamarckian (i.e. inheritance of acquired characteristics) dimension to the evolution of the vertebrate immune system. Controversial though the proposition of such a Lamarckian mechanism may be, recent molecular studies of epigenetic regulation indicate that it would not be unprecedented (Balter, 2000). In any case, there can be little doubt that the emergence of specific immunity had a deep impact on vertebrate evolution. The probable role of transposable elements in this process underscores their ability to influence the evolution of increasingly complex cellular machinery.

Cancer

While not directly related to the evolution of organismic complexity, it may be interesting to speculate as to the relationship between cancer and transposable elements. Cancer cells, replicating unchecked and out of control can be considered to be entities that have fallen into an unordered chaotic regulatory regime (Pienta *et al*., 1989; Coffey, 1998; Fedoroff, 1999). In many instances, part of this regulatory breakdown results in the increased expression of transposable elements including many endogenous retroviruses (Bieda *et al*., 2001; Kim *et al*., 2001; Wang-Johanning *et al*., 2001). This is likely to be due to changes in patterns of DNA methylation. A large body of data links aberrant DNA methylation to the inactivation of tumor suppressor genes (Esteller and Herman, 2002). In a non-cancerous cell the majority of DNA methylation is associated with silent transposable element sequences (Wade, 2001). In tumor cells, a global

hypomethylation occurs followed by hypermethylation of many of the CpG islands found in the regulatory regions of cellular genes. This redistribution of CpG methylation leads to the simultaneous destructive inactivation of tumor suppressor genes and transcriptional activation of transposable element sequences. What is unknown is how and why the methylation leaves the transposable elements and is then transferred to host genes. It seems probable that transposable elements play some role in initiating the regulatory breakdown that results in cancer. However, the focus on understanding methylation changes in cancer remains centered on host genes. Given the known functions of host tumor suppressor genes this is understandable, but a more nuanced understanding of the relationship between methylation and cancer would seem to require some knowledge of the role that elements play in the process. For example, it may be the case that the inactivation of any given tumor supressor gene is merely epiphenomenal. In other words, it is not the inactivation of the host gene that leads to cancer but rather the activation of the unmethylated elements. Transposable element activation could cause enhanced levels of transposition that could in turn cause the types of chromosomal re-arrangements common in many different cancers. In addition, DNA methylation is known to repress recombination. Another way element hypomethylation could lead to chromosomal re-arrangements and cancer is through ectopic recombination between newly unmethylated element sequences. The molecular phenomenology of cancer is still largely a mystery. A focus on the role of transposable elements in this process would be novel and could help to provide a global genomic perspective on the changes that take place during cancer.

Conclusion

The obvious increase in organismic complexity that has occurred over evolutionary time begs the question of whether or not this increase represents an active and directional trend. When the totality of organic life is considered, this increase does in fact appear to be a passive and random result of an increase in the total variation of complexity. However, there is evidence that directed complexity increases do occur within related groups of organisms (clades). In a very broad sense, eukaryotes represent a single organismic clade. The vast majority of eukaryotic phylogenetic diversity is found among single celled protozoans. Just as with the bacteria and archaea, the dominant mode of eukaryotic life can be considered to be relative simple and single-celled, and so the overall increase of eukaryotic complexity may also be passive and random. On the other hand, the less phylogentically diverse eukaryotic crown group contains a plethora of complex multi-cellular organisms that represents an abundance of organismic diversity. Thus the complexity increase in the crown group of organisms that evolved from single celled ancestors can be considered to have been an active and directed one.

Ubiquitous and abundant, transposable elements have been major players in eukaryotic genome evolution. Host-element evolutionary dynamics often resemble those of host and parasite as well as the arms races seen for predator-prey relationships. The intragenomic conflicts that these host-element relationships entail could well have supplied a motive force to help bring about an active and directional trend of increasing eukaryotic complexity. Consistent with this idea, consideration of the numerous ways that transposable elements have influenced organismic evolution indicates that they have in fact had a hand in driving the development of a number of cellular systems that are hallmarks of eukaryotic complexity.

Acknowledgements

We would like to thank Fyodor A. Kondrashov for discussion on the evolution of sexual reproduction and John F. McDonald for inspiration and guidance.

References

Agrawal, A., Eastman, Q.M., and Schatz, D.G. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature. 394: 744-751.

Balter, M. 2000. Genetics. Was Lamarck just a little bit right? Science. 288: 38.

Barton, N.H., and Charlesworth, B. 1998. Why sex and recombination? Science. 281: 1986-1990.

Behrens, R., Hayles, J., and Nurse, P. 2000. Fission yeast retrotransposon Tf1 integration is targeted to 5' ends of open reading frames. Nucleic Acids Res. 28: 4709-4716.

Berezney, R., and Wei, X. 1998. The new paradigm: integrating genomic function and nuclear architecture. J. Cell. Biochem. Suppl. 31: 238-242.

Bestor, T.H. 1999. Sex brings transposons and genomes into conflict. Genetica. 107: 289-295.

Bieda, K., Hoffmann, A., and Boller, K. 2001. Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines. J. Gen. Virol. 82: 591-596.

Bird, A.P. 1995. Gene number, noise reduction and biological complexity. Trends Genet. 11: 94-100.

Blanden, R.V., Rothenfluh, H.S., Zylstra, P., Weiller, G.F., and Steele, E.J. 1998a. The signature of somatic hypermutation appears to be written into the germline IgV segment repertoire. Immunol. Rev. 162: 117-132.

Blanden, R.V., Rothenfluth, H.S., and Steele, E.J. 1998b. On the possible role of natural reverse genetics in the V gene loci. Curr. Top. Microbiol. Immunol. 229: 21-32.

Blanden, R.V., and Steele, E.J. 1998. A unifying hypothesis for the molecular mechanism of somatic mutation and gene conversion in rearranged immunoglobulin variable genes. Immunol. Cell. Biol. 76: 288-293.

Britten, R.J. 1996. DNA sequence insertion and evolutionary variation in gene regulation. Proc. Natl. Acad. Sci. USA. 93: 9374-9377.

Britten, R.J., and Davidson, E.H. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. 46: 111-138.

Brosius, J. 1999. Genomes were forged by massive bombardments with retroelements and retrosequences. Genetica. 107: 209-238.

Carroll, S.B. 2001. Chance and necessity: the evolution of morphological complexity and diversity. Nature. 409: 1102-1109.

Cavalier-Smith, T., and Beaton, M.J. 1999. The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. Genetica. 106: 3-13.

Charlesworth, B., Sniegowski, P., and Stephan, W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. 371: 215-220.

Coffey, D.S. 1998. Self-organization, complexity and chaos: the new biology for medicine. Nat. Med. 4: 882-885.

Consortium, T.C.e.S. 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. The *C. elegans* Sequencing Consortium. Science. 282: 2012-2018.

Doolittle, W.F., and Sapienza, C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature. 284: 601-603.

Dorer, D.R., and Henikoff, S. 1994. Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. Cell. 77: 993-1002.

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.T., Wu, X., Reith, M., Cavalier-Smith, T., and Maier, U.G. 2001. The highly reduced genome of an enslaved algal nucleus. Nature. 410: 1091-1096.

Esteller, M., and Herman, J.G. 2002. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. J. Pathol. 196: 1-7.

Fedoroff, N.V. 1999. Transposable elements as a molecular evolutionary force. Ann. NY. Acad. Sci. 870: 251-264.

Finnegan, D.J. 1992. Transposable elements. Curr. Opin. Genet. Dev. 2: 861-867.

Fisher, R.A. 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Futuyma, D.J. 1986. Evolutionary Biology. Sinauer Associates, Sunderland, Massachusetts.

Gerasimova, T.I., Byrd, K., and Corces, V.G. 2000. A chromatin insulator determines the nuclear localization of DNA. Mol. Cell. 6: 1025-1035.

Gould, S.J. 1996. Full House: The Spread of Excellence from Plato to Darwin. Harmony Books, New York, New York.

Hickey, D.A. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics. 101: 519-531.

Hiom, K., Melek, M., and Gellert, M. 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell. 94: 463-470.

Jacob, F. 1977. Evolution and tinkering. Science. 196: 1161-1166.

Jensen, S., Gassama, M.P., and Heidmann, T. 1999. Taming of transposable elements by homology-dependent gene silencing. Nat. Genet. 21: 209-212.

Jordan, M.A., Saedler, H., and Starlinger, P. 1967. Strong-polar mutations in the transferase gene of the galactose operon. E. coli Mol. Gen. Genet. 100: 296-306.

Kidwell, M.G., and Lisch, D. 1997. Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. USA. 94: 7704-7711.

Kidwell, M.G., and Lisch, D.R. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution. 55: 1-24.

Kim, H.S., Yi, J.M., and Jeon, S.H. 2001. Isolation and phylogenetic analysis of HERV-K long terminal repeat cDNA in cancer cells. AIDS Res. Hum. Retroviruses. 17: 987-990.

Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A., and Voytas, D.F. 1998. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res. 8: 464-478.

Kondrashov, A.S. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature. 336: 435-440.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. Nature. 409: 860-921.

Leeton, P.R., and Smyth, D.R. 1993. An abundant LINE-like element amplified in the genome of Lilium speciosum. Mol. Gen. Genet. 237: 97-104.

Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., and Sheen, F.M. 1993. Transposons in place of telomeric repeats at a Drosophila telomere. Cell. 75: 1083-1093.

Lewis, S.M., and Wu, G.E. 1997. The origins of V(D)J recombination. Cell. 88: 159-162.

Li, W.H., and Graur, D. 1991. Fundamentals of Molecular Evolution. Sinauer Associates, Sunderland, Massachusetts.

Li, W.H., Gu, Z., Wang, H., and Nekrutenko, A. 2001. Evolutionary analyses of the human genome. Nature. 409: 847-849.

Lohe, A.R., Moriyama, E.N., Lidholm, D.A., and Hartl, D.L. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. Mol. Biol. Evol. 12: 62-72.

Makalowski, W. 2000. Genomic scrap yard: how genomes utilize all that junk. Gene. 259: 61-67.

Martienssen, R. 1996. Epigenetic phenomena: paramutation and gene silencing in plants. Curr. Biol. 6: 810-813.

Matzke, M.A., Mette, M.F., Aufsatz, W., Jakowitsch, J., and Matzke, A.J. 1999. Host defenses to parasitic sequences and the evolution of epigenetic control mechanisms. Genetica. 107: 271-287.

Maynard-Smith, J. 1978. The evolution of sex. Cambridge Univeristy Press. Cambridge.

McClintock, B. 1948. Mutable loci in maize. Carnegie Institution of Washington Year Book. 47: 155-169.

McDonald, J.F. 1993. Evolution and consequences of transposable elements. Curr. Opin. Genet. Dev. 3: 855-864.

McDonald, J.F. 1998. Transposable elements, gene silencing and macroevolution. Trends Ecol. Evol. 13: 94-95.

Mighell, A.J., Markham, A.F., and Robinson, P.A. 1997. Alu sequences. FEBS Lett. 417: 1-5.

Miller, W.J., McDonald, J.F., and Pinsker, W. 1997. Molecular domestication of mobile elements. Genetica. 100: 261-270.

Muller, H.J. 1932. Some genetic aspects of sex. Amer. Nat. 66: 118-138.

Nabirochkin, S., Ossokina, M., and Heidmann, T. 1998. A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. J. Biol. Chem. 273: 2473-2479.

Nekrutenko, A., and Li, W.H. 2001. Transposable elements are found in a large number of human protein- coding genes. Trends Genet. 17: 619-621.

Orgel, L.E., and Crick, F.H. 1980. Selfish DNA: the ultimate parasite. Nature. 284: 604-607.

Pardue, M.L., and DeBaryshe, P.G. 1999. Drosophila telomeres: two transposable elements with important roles in chromosomes. Genetica. 107: 189-196.

Pienta, K.J., Partin, A.W., and Coffey, D.S. 1989. Cancer as a disease of DNA organization and dynamic cell structure. Cancer Res. 49: 2525-2532.

Rio, D.C. 1990. Molecular mechanisms regulating Drosophila P element transposition. Annu. Rev. Genet. 24: 543-578.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science. 274: 765-768.

Sheen, F.M., and Levis, R.W. 1994. Transposition of the LINE-like retrotransposon TART to Drosophila chromosome termini. Proc. Natl. Acad. Sci. USA. 91: 12510-12514.

Spanopoulou, E., Zaitseva, F., Wang, F.H., Santagata, S., Baltimore, D., and Panayotou, G. 1996. The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. Cell. 87: 263-276.

Stavenhagen, J.B., and Robins, D.M. 1988. An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. Cell. 55: 247-254.

Steele, E.J., Robyn, R.A., and Blanden, R.V. 1998. Lamarcks Signature: How Retrogenes are Changing Darwin's Natural Selection Paradigm. Perseus Books, Reading, Massachusetts.

Steele, E.J., Rothenfluh, H.S., and Blanden, R.V. 1997. Mechanism of antigen-driven somatic hypermutation of rearranged immunoglobulin V(D)J genes in the mouse. Immunol. Cell. Biol. 75: 82-95.

van den Berg, C.W., Demant, P., Aerts, P.C., and Van Dijk, H. 1992. Slp is an essential component of an EDTA-resistant activation pathway of mouse complement. Proc. Natl. Acad. Sci. USA. 89: 10711-10715.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. Science. 291: 1304-1351.

Wade, P.A. 2001. Methyl CpG-binding proteins and transcriptional repression. Bioessays. 23: 1131-1137.

Wang-Johanning, F., Frost, A.R., Johanning, G.L., Khazaeli, M.B., LoBuglio, A.F., Shaw, D.R., and Strong, T.V. 2001. Expression of human endogenous retrovirus k envelope transcripts in human breast cancer. Clin. Cancer Res. 7: 1553-1560.

Wessler, S.R., Bureau, T.E., and White, S.E. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. 5: 814-821.

White, S.E., Habera, L.F., and Wessler, S.R. 1994. Retrotransposons in the flanking regions of normal plant genes: a role for copia-like elements in the evolution of gene structure and expression. Proc. Natl. Acad. Sci. USA. 91: 11792-11796.

Wolffe, A.P., and Matzke, M.A. 1999. Epigenetics: regulation through repression. Science. 286: 481-486.

Wright, R. 2000. Non Zero: the Logic of Human Destiny. Vintage Books, New York, New York.

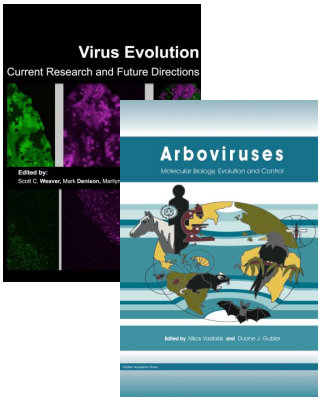
Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. 2001. Genome-Wide Distribution of ORC and MCM Proteins in S. cerevisiae: High- Resolution Mapping of Replication Origins. Science. 294: 2357-2360.

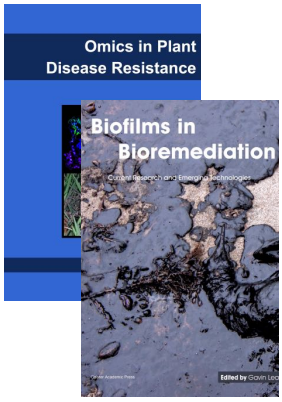
Yoder, J.A., Walsh, C.P., and Bestor, T.H. 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13: 335-340.

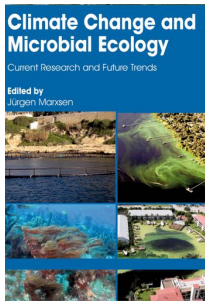
Further Reading

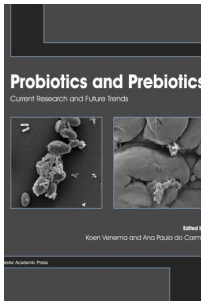
Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at **caister.com**

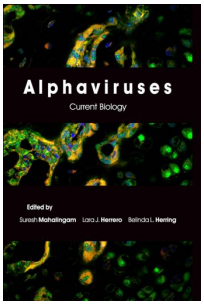
- **MALDI-TOF Mass Spectrometry in Microbiology**
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof
- ***Aspergillus* and *Penicillium* in the Post-genomic Era**
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2
- **The Bacteriocins: Current Knowledge and Future Prospects**
  Edited by: RL Dorit, SM Roy, MA Riley (2016)
  www.caister.com/bacteriocins
- **Omics in Plant Disease Resistance**
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr
- **Acidophiles: Life in Extremely Acidic Environments**
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles
- **Climate Change and Microbial Ecology: Current Research and Future Trends**
  Edited by: J Marxsen (2016)
  www.caister.com/climate
- **Biofilms in Bioremediation: Current Research and Emerging Technologies**
  Edited by: G Lear (2016)
  www.caister.com/biorem
- **Microalgae: Current Research and Applications**
  Edited by: MN Tsaloglou (2016)
  www.caister.com/microalgae
- **Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives**
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma
- **Virus Evolution: Current Research and Future Directions**
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusevol
- **Arboviruses: Molecular Biology, Evolution and Control**
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo
- ***Shigella*: Molecular and Cellular Biology**
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella
- **Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment**
  Edited by: AM Romaní, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms
- **Alphaviruses: Current Biology**
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha
- **Thermophilic Microorganisms**
  Edited by: F Li (2015)
  www.caister.com/thermophile



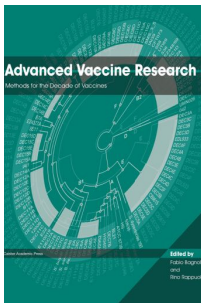


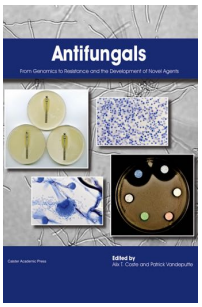












- **Flow Cytometry in Microbiology: Technology and Applications**
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow
- **Probiotics and Prebiotics: Current Research and Future Trends**
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics
- **Epigenetics: Current Research and Emerging Trends**
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics2015
- ***Corynebacterium glutamicum*: From Systems Biology to Biotechnological Applications**
  Edited by: A Burkovski (2015)
  www.caister.com/cory2
- **Advanced Vaccine Research Methods for the Decade of Vaccines**
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines
- **Antifungals: From Genomics to Resistance and the Development of Novel Agents**
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals
- **Bacteria-Plant Interactions: Advanced Research and Future Trends**
  Edited by: J Murillo, BA Vinatzer, RW Jackson, et al. (2015)
  www.caister.com/bacteria-plant
- ***Aeromonas***
  Edited by: J Graf (2015)
  www.caister.com/aeromonas
- **Antibiotics: Current Innovations and Future Trends**
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics
- ***Leishmania*: Current Biology and Control**
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2
- ***Acanthamoeba*: Biology and Pathogenesis (2nd edition)**
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2
- **Microarrays: Current Technology, Innovations and Applications**
  Edited by: Z He (2014)
  www.caister.com/microarrays2
- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order