Genetica **115:** 49–63, 2002.

Transposable elements and the evolution of genome size in eukaryotes

Margaret G. Kidwell
Department of Ecology and Evolutionary Biology, The University of Arizona, Tucson, AZ 85721, USA (Phone: (520) 621-1784; Fax: (520) 621-9190; E-mail: kidwell@azstarnet.com)

Key words: genome size, molecular evolution, transposable element

Abstract

It is generally accepted that the wide variation in genome size observed among eukaryotic species is more closely correlated with the amount of repetitive DNA than with the number of coding genes. Major types of repetitive DNA include transposable elements, satellite DNAs, simple sequences and tandem repeats, but reliable estimates of the relative contributions of these various types to total genome size have been hard to obtain. With the advent of genome sequencing, such information is starting to become available, but no firm conclusions can yet be made from the limited data currently available. Here, the ways in which transposable elements contribute both directly and indirectly to genome size variation are explored. Limited evidence is provided to support the existence of an approximately linear relationship between total transposable element DNA and genome size. Copy numbers per family are low and globally constrained in small genomes, but vary widely in large genomes. Thus, the partial release of transposable element copy number constraints appears to be a major characteristic of large genomes.

Introduction

In contrast to the relatively narrow range of genome sizes in prokaryotes, eukaryotic species vary more than 200,000-fold in the size of their genomes (Gregory, 2001). No single factor, such as number of coding genes or degree of organismal complexity is closely correlated with genome size, giving rise to the so-called 'C-value paradox'. Today, it is generally accepted that differential amounts of non-coding, repetitive, DNA account for a major fraction of eukaryotic genome size variation, thus providing a partial explanation of the paradox (Gregory & Hebert, 1999; Petrov, 2001). However, a whole set of questions remain unanswered, including the relative contributions of various types of repetitive DNA, such as transposable elements, satellite DNAs and simple sequence repeats to the variation in genome size. It is also unclear how genome sizes are contained or reduced. These questions are highly complex with no simple answers likely to be found because the evolutionary forces shaping genome size vary widely among different organisms (Petrov, 2001).

Bearing in mind that genome sequence data are mostly incomplete and limited to a few species, the purpose of this paper is to initiate a broad scale survey, involving both animals and plants, into the extent to which transposable elements determine genome size. A major aim will be to inquire whether or not there is evidence for any broad generalizations concerning the types and copy numbers of transposable elements that distinguish small from larger genomes.

Transposable elements and genome evolution

Transposable elements are potent mutagenic agents with the potential to produce a wide array of changes in the genomes of their hosts (Kidwell & Lisch, 2000; 2001). The range of 'mutations' induced by transposable element activity extends from modifications in the size and arrangement of whole genomes to substitutions, deletions, and insertions of a single nucleotide. In addition to their ability to increase genome size following transposition, transposable elements induce chromosomal rearrangements such as deletions,

duplications, inversions, and reciprocal translocations. This activity provides the potential for small- and large-scale genome reorganization, amplification, and reduction.

Transposable elements have been found at the origin of numerous types of chromosomal rearrangements in many species (e.g., Lim & Simmons, 1994; Caceres et al., 1999) Increases in transposable element copy number following transposition have also been directly implicated in significant genome size expansions over a relatively short period of evolutionary time (e.g., a doubling of the maize genome during the last few million years (SanMiguel et al., 1998)). Transposable elements are undoubtedly responsible for a significant proportion of the observed karyotypic variation among many groups. Reliable estimates of the fraction of all rearrangements accounted for by transposable elements are currently not available in any species, but are urgently needed. These elements may also produce mutations when they excise imprecisely, leaving either no identifying sequence, or only small 'footprints' of their previous presence. A major problem of identification arises because of rapid divergence of nonautonomous elements and the origin of many ancient sequences may be difficult, if not impossible, to identify.

Gradual changes versus quantum jumps

A useful distinction has been made between global and local forces that act to modulate genome size (Petrov, 2001). It is expected that relatively fast genome expansion will be modulated by the action of global forces. Both global and local forces would be expected to contribute to more gradual changes. Transposable elements, along with polyploidization, are good examples of global forces for expansion, whereas expansion of microsatellites, heterochromatin, or various types of tandemly repeated sequences, represent local forces.

Several interesting patterns of genome size variability have been observed. For example, among birds, mammals and teleost fish, genome size variation is relatively small, suggesting that changes occur by gradual accretion of small blocks of DNA (Gregory & Hebert, 1999). However, the large variation in genome size observed among invertebrates and plants suggest that quantum jumps make an important contribution and this has been well documented in some plants (e.g., maize SanMiguel et al., 1996, 1998). It will be interesting to determine to what extent transposition is associated with one or both of these patterns of genome size increase.

Transposable element classification

Most transposable elements can be assigned to two main classes (see Table 1), according to their mechanism of transposition (Finnegan, 1989; Capy et al., 1997). Both autonomous and nonautonomous members are found in many element families of both classes. Autonomous elements are able to catalyze their own transposition, while nonautonomous, elements depend on autonomous elements from the same family for transposition. Class I elements are members of the larger group of Retroid agents that also includes retroviruses (McClure, 1999). They use an RNA-mediated mode of transposition and encode a reverse transcriptase (RT). These elements are amplified by the transposition process and have a high potential for increase in copy number. They are divided into two subclasses, the retrotransposons that are characterized by direct, long terminal repeats (LTRs), and the retroposons that lack terminal repeats. Here these elements will be referred to as LTR retroelements (or LTR retrotransposons) and non-LTR retroelements, respectively. The LTR retroelements include four distinct groups (Malik et al., 2000), *Ty1-copia, BEL, DIRS*, and *Ty-3 gypsy,* that are broadly distributed in many animals and plants. The non-LTR retroelements include the long interspersed nuclear elements (*LINE*s) and the short interspersed nuclear elements (*SINE*s).

Many Class II elements, the transposons (*sensu strictu*), use a DNA-mediated mode of 'cut and paste' transposition. However, it was recently shown that a number of eukaryotic Class II elements, called *Helitrons*, have probably propagated by a mechanism similar to rolling-circle (RC) transposition in prokaryotes (Kapitonov & Jurka, 2001). *Helitrons* tend to be large (Table 1) and, surprisingly, constitute as much as 2% of the genome of *Arabidopsis thaliana* and *Caenorhabditis elegans* (Kapitonov & Jurka, 2001). The 'cut and paste' Class II elements include the *hAT* (*hobo, Activator, Tam-3*) superfamily of elements, the *mariner*-like superfamily of elements, the *P* elements, and the *MuDR* elements. Although the miniature inverted repeat transposable elements (*MITE*s) and the *foldback* (*FB*) elements were earlier assigned to a third class (Capy et al., 1997), subsequent evidence strongly suggests that *MITE*s are members of the Class II elements (Le et al., 2000; Turcotte et al., 2001). In

Table 1. Characteristics of some widely distributed types of transposable elements

Class	Subclass	Superfamily	Family examples	Approximate size range (bp)
I. Retroelements RNA-mediated elements	LTR retro-transposons	*Ty1-copia*	*Opie-1* in maize	3000–12,000
		BEL	*Cer7* in *C. elegans*	3000–20,000
		DIRS-1	*DIRS-1* in *Dictyostelium*	∼5000
		Ty3-gypsy	*Gypsy* in *Drosophila*	5000–14,000
	Non-LTR retroposons	*LINE*s	*LINE-1* in humans; *I* element in *Drosophila*	1000–7000
		*SINE*s	*Alu* in humans	100–500
II. Transposons DNA-mediated elements	Cut and paste transposition DDE signature present	*mariner-Tc1*	*Tc1* in *C. elegans*; *mariner* in *Drosophila*	1000–2000
		Mu	*Mu* in maize; *MULE*s in *Arabidopsis*	400–20,000
		*MITE*s	*Tourist* in maize	100–500
	Cut and paste transposition DDE signature absent	*hAT*	*hobo* in *Drosophila;* *Ac in* maize; *Tam-3* in *Anthirinhum*	500–4600
		P	*P* in *Drosophila*	
	Rolling circle (RC) transposition	*Helitrons*	*Helitrons* in *A. thaliana*, *O. sativa*, and *C. elegans*	5500–17,500
Unclassified		*Foldback*	*Galileo* in *D. buzzatii*	Large, but highly variable
		Mini-me	*Mini-me* in many *Drosophila* species	500–1200

general, the potential for copy number increase appears to be more restricted for Class II than for Class I elements.

As shown in Table 1, different superfamilies and families of transposable elements have widely differing characteristic sizes. *Helitrons* and LTR retrotransposons tend to be large. Lengths from 3 to more than 17 kb, or more, have been observed for full-sized elements. The size of non-LTR retroelements varies from several kilobases (typical of *LINE*s) to <500 bp, typical of *SINE*s. *MITE*s are also relatively small (140–500 bp). Even smaller are *Maque* elements in *Anopheles gambiae* whose size is a minute 60 bp (Tu, 2001b). These elements have been hypothesized to result from incomplete reverse transcription of non-LTR retrotransposons and may represent the progenitors of *SINE*s (Tu, 2001b). *Mutator*-like elements, or *MULE*s, represent a family of transposable

elements whose lengths are unusually variable. For example, the size range of individual *MULE*s in *A. thaliana* is from 444 to 19,397 bp (Yu et al., 2000).

The difference in size between large and small transposable elements has obvious implications for their relative contributions to genome size variation. As the contribution of transposable elements to genome size is a product of copy number and element length, fewer numbers of large elements will be needed to make a contribution equivalent to a larger number of smaller elements. This is well illustrated in the slime mold genome in which non-LTR retroelements and DNA elements are each represented by a total of 235 transposable element elements in seven families (Glockner et al., 2001). However, because of their relatively larger average size, the non-LTR elements contribute a total of 3.7% to the genome, compared with only 1.5% for the equivalent number of DNA elements.

A characteristic that is common to both Class I and II elements, and which needs emphasis, is that fully functional elements tend to be in the minority in most genomes. The vast majority of copies are usually nonautonomous elements, or even small fragments of full-size copies.

Target site preference and transposable element copy number

Broad patterns of non-random transposable element distribution have been recorded in many organisms, often related to the type of element involved (reviewed by Kidwell & Lisch, 2001). Some transposable elements preferentially insert into regions away from host gene sequences, such as intergenic regions, heterochromatin, or, frequently into other transposable elements. For example, in plants, the most abundant LTR retrotransposons, sometimes described as 'intergenic LTR retrotransposons', are found most frequently in methylated, presumably locally heterochromatic regions, often in nested clusters (SanMiguel et al., 1996). In maize, these elements strongly prefer to insert into the LTRs of similar elements (Bennetzen, 2000). Such elements appear to have relatively low constraints on copy number increase. Other transposable elements insert preferentially into, or near, single copy sequences. For example, in plants, some DNA transposable elements exhibit preferential insertion and/or retention within euchromatic regions of the genome that are genetically active and unmethylated (Cresse et al., 1995; Bennetzen, 2000; Zhang et al., 2000). *Mutator* (*Mu*) elements in maize specifically target gene sequences and low copy number DNA (Cresse et al., 1995). Some LTR retrotransposons in plants also share this pattern (Garber et al., 1999). Perhaps the low copy numbers of this type of element are due to greater selection against unrestricted transposition in gene-rich regions (Kidwell & Lisch, 2001).

The relationship between transposable elements and genome size

With the recognition that repetitive DNA accounts for a large fraction of the genome in most organisms, a number of authors have speculated that transposable elements play an important role in accounting for the C-value paradox, particularly in plants. For example, it has been claimed that transposable elements make up the major type of identified non-genic DNA in every plant species (Bennetzen, 2000). In this case, non-genic DNA probably means non-coding DNA and includes introns. Furthermore, an important role has been claimed for LTR retrotransposons in determining the size of plant genomes, in general (Kumar, 1996; Kumar & Bennetzen, 1999). However, in the opinion of Wendel and Wessler (2000), while differential amplification of retrotransposons largely accounts for the C-value paradox in grass genomes, factors other than transposable elements appear to have a greater influence on genome size differences in other organisms. It will be interesting to see whether sequence data support this distinction between grasses and other plant genomes.

Comparative studies of animal genomes will be important in determining the contributions of transposable elements, as well as other types of repetitive DNA, to genome size variation. For example, much of the human Y chromosome is composed of repetitive satellite DNA, in addition to being rich in *LINE*s and *SINE*s (e.g., Tilford et al., 2001). In contrast to the paucity of data for many other animal taxa, genome size has been extensively studied in some mosquitoes, using quantitative cytophotometry and analysis of reassociation kinetics of nuclear DNA. Overall, a general increase in genome size has been identified in the evolution of the Culicidae (Rai & Black, 1999). Earlier, Black and Rai (1988) demonstrated that all classes of repetitive DNA sequences increased linearly in amount with total genome size.

With the limited information currently available, it is not easy to tease out the contributions of transposable elements from that of other types of repetitive DNA in the mosquito species mentioned above (Rai & Black, 1999). However, others have recently studied *MITE*s in several mosquito species. In *Culex pipiens* (540 Mbp) the *Mimo* family is present in ca. 1000 copies (Feschotte & Mouches, 2000). This is higher than the average copy number (range 40–1340 copies) reported for eight families of *MITE*s in *An. gambiae* (270 Mbp) (Tu, 2001a), but lower than that of 2100–10,000 copies found in *Aedes aegypti* (810 Mbp) (Tu, 1997, 2000). It is consistent with the hypothesis that *MITE* copy number may be correlated with host genome size (Tu, 1997).

We now examine the limited information currently available from recent gene and genome sequencing projects on the contributions of transposable elements to genome size, first in terms of direct measurable contributions and, second, in terms of indirect, quantitatively unmeasurable contributions.

Direct contributions of transposable elements to interspecific genome size variation

The question of how much of the interspecific variation in genome size can be accounted for by transposable elements is, in principal, a straightforward one. In practice, although estimates of genome size are available for many different species, estimates of the size of transposable element genomic complements are currently available for only a very few, well-studied, species. Further, at best, these often provide only very crude estimates of element numbers in non-euchromatic regions of genomes. This situation is expected to improve with the increasing number of genome sequencing projects that are ongoing. However, because of the difficulties inherent in sequencing repetitive DNA, this fraction of the genome is generally assigned a low priority and reliable data are often slow to be published. Also, given the large variation in genome size within at least some species, even sequencing projects may provide only partial information.

In Table 2, data are presented on genomic transposable element fractions for 12 species, including humans and a number of model organisms. The proportion of bulk DNA contributed by transposable elements to this sample of genomes varies widely from 2% in the puffer fish, *Fugu rubripes*, to >50% of the genomes of the cereal grasses, maize, and barley. Estimates for the total DNA contributed by transposable elements to these 12 genomes was plotted against genome size (Figure 1). Overall, the linear regression equation $Y = -92.41 + 0.51715X$ provided a good fit to the data ($t = 6.43$ for 10 d.f.; $p \leqslant 0.001$). Although an S-shaped logistic curve did not provide any better fit to the data than the linear regression equation, visual inspection of Figure 1 suggests that factors other than transposable elements may contribute more to genome size variation in the smaller than in the larger genomes. Also, it seems likely that the true relationship between transposable elements and genome size is not a simple linear one. An exponential rate of increase might be expected for elements that insert into their own kind, but not for others that insert directly into host DNA. More detailed models of the various processes involved will be required to better understand what is expected to be a complex relationship.

On the basis of this preliminary analysis the following working hypothesis is proposed for testing in future work: The contribution of transposable elements to genome size variation is greater, relative to other sources of variation, in larger (>500 Mbp) than in smaller genomes (<500 Mbp). (Note the definition of small and large genomes is quite arbitrary at this point and was dictated by the genome sizes of the species available). Therefore, as has been suggested earlier, transposable elements may play a more important role in the increase in size of relatively large plant and animal genomes than in smaller ones. However, plant transposable elements may vary from those of animals in the details of their effects on genome size expansion, as discussed below.

Transposable element distribution patterns in small and large genomes

In the last section, we proposed that transposable elements are relatively more important determinants of genome size in large compared with small genomes. We now ask how small and large genomes differ in their transposable element distributions. Is the higher proportion of transposable elements found in large genomes accounted for by an increase in number of transposable element families? Additionally, or alternatively, is it explained by an increase in copy number per family? Are there specific transposable element superfamilies or families that are better represented in

Table 2. Genome sizes and transposable element (TE) proportions for 12 species

Species	Genome	No. coding	% TE	References
S. cerevisiae	12		3	(Kim et al., 1998)
D. discoideum	34		10	(Glockner et al., 2001)
C. elegans	100	18,400	6	(Waterston & Sulston, 1995; International Human Genome Sequencing Consortium 2001)
A. thaliana	125	25,498	14	(The Arabidopsis Genome Initiative, 2000)
D. melanogaster	180	13,600	15	(Vieira et al., 1999; Adams et al., 2000)
A. quadrimaculatus	245		16	(Rai & Black, 1999)
F. rubripes	400		2	(Elgar et al., 1999)
O. sativa	430		14	(Turcotte et al., 2001)
Z. mays	2500	50,000	60	(SanMiguel et al., 1996)
H. sapiens	3000	30,000	44	(International Human Genome Sequencing Consortium, 2001).
M. musculus	3250		40	(Henikoff et al., 1997; Smit, 1999)
H. vulgare	5000		55	(Kumar & Bennetzen, 1999; Vicient et al., 1999)

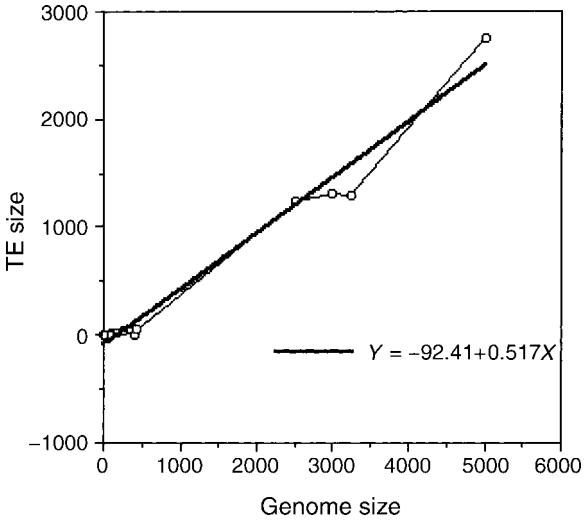

Figure 1. The relationship between the total amount of transposable element DNA (TE size) and genome size for the 12 species listed in Table 2 is plotted along with the linear regression equation $Y = -92.41 + 0.51715X$.

large genomes relative to small ones? There are conflicting opinions about the answers to these questions suggesting that more detailed examination is needed. For example, in the published report of the initial sequencing of the human genome (International Human Genome Sequencing Consortium, 2001), it was stated 'the worm, the fly, and mustard weed genomes all contain many transposon families, each consisting of hundreds to thousands of elements'. The source of evidence for this statement was not provided, but it suggests that, contrary to many previous published reports, transposable element copy number in small genomes is very large.

In plants, it has been claimed that genome size variation is correlated with both number of retrotransposon families and total copy number (Kumar & Bennetzen, 1999). However, it is not clear if this generalization holds for types of transposable elements other than retrotransposons in plants, and for all elements in animals. Accordingly, we now look at available data to try to begin to answer these questions.

Sources of transposable element copy number data

As seen in Table 3, comparative data on transposable element copy numbers were obtained from the literature for the following five eukaryotic species having small genomes: *Saccharomyces cerevisiae* (yeast), *Dictvostelium discoideum* (slime mold), *A. thaliana* (mustard weed), *C. elegans* (worm), and *D. melanogaster* (fly). It should be emphasized that these data are very preliminary because, of these five, genome

Table 3. Transposable element copy numbers in five small genomes compared with that of humans

	Yeast	Slime mold	Worm	Mustard weed	Fly	Human
Genome size (Mb)	12	34	100	125	180	3400
LTR						
No. copies	331	325	24	1594	317	443×10^3
No. families	5	6	1	70*	22	104
Mean copy no. (SE)	66.2 (± 38.1)	54.2 (± 36.8)	24.0	22.8	14.5 (± 8.0)	4259
% of genome	3.1	4.4	0.1	6.4		8.3
Non-LTR retroelements						
No. copies	0	235	611	515	87	2426×10^3
No. families	0	7	12	10*	5	6
Mean copy no. (SE)	0	33.6 (± 10.1)	50.9 (± 8.7)	51.5	17.3 (± 6.4)	4043×10^2
% of genome	0	3.7	0.4*	0.7		33.6
DNA elements						
No. copies	0	235	3083	2203	82	294×10^3
No. families	0	7	12	80*	4	63
Mean copy no. (SE)	0	33.6 (±16.6)	256.9 (± 89)	27.5	20.5 (± 10.1)	4667
% of genome	0	1.5	5.3*	6.8		2.8
Total elements						
No. copies	331	795	3718	5602	486	3163×10^3
No. families	5	20	25	180*	31	263
Mean copy no. (SE)	66.2 (± 38.1)	39.8 (± 12.4)	148.7 (± 47.0)	31.1	15.7 (± 2.7)	12,049.4
% of genome	3.1	9.6	6.5*	14	10–12	44.8
References	(Kim et al., 1998)	(Glockner et al., 2001)	(Duret et al. 2000; *International Human Genome Sequencing Consortium, 2001)	(The Arabidopsis Genome Initiative, 2000; *International Human Genome Sequencing Consortium, 2001)	(Vieira et al., 1999)	(International Human Genome Sequencing Consortium, 2001)

sequencing is only complete for yeast. Representing large genomes, the first draft of the human genome sequence (ca. 3000 Mbp) was included for comparison. For each of the three major categories, LTR retroelements, non-LTR retroelements, and DNA elements, it was of interest to ascertain the approximate number of transposable element families and their mean copy numbers. However, the limitations in the published data make this data very preliminary. For example, in many cases there is no general agreement about the definitions of families and subfamilies. Also, copy numbers reported for some families contain only full-sized elements, while in other cases deletion derivatives are included.

Transposable element comparisons among five small genomes

Sizes of the five small genomes included in Table 3 range from 12 Mbp for yeast to 180 Mbp for fly. Yeast transposable elements are exceptional in being represented by only five families of LTR retroelements. No members of the non-LTR retroelements, or DNA elements, are found in this species. In contrast, all three major element categories are represented in the other four small genomes.

The 20 families of transposable elements found in slime mold are fairly evenly divided between LTR, non-LTR, and DNA elements (Table 3). In the worm,

LTR retroelements are represented by only a single family and mean copy number per family is five times larger for DNA elements than for non-LTR retroelements. While both LTR and DNA elements predominate in mustard weed, the LTR retroelements in the fly predominate over those of the other two major categories. The number of families per transposable element category varied, but was in no case greater than 100. Mean copy number per family was modest and only exceeded 100 for DNA elements in the worm.

In summary, for the five small genomes, the number of transposable element families is quite variable. However, contrary to the earlier claim (International Human Genome Sequencing Consortium, 2001), copy numbers per family are small. Moreover, the relative representation of the three major element categories is highly variable, indicating no clear predominance of any one category of transposable element in these genomes.

Comparison with transposable elements in the human genome

The most striking difference between the transposable element characteristics of the human genome compared with those of the five small genomes (Table 3) is the major increase in copy number per family in humans. This difference is greatest for the six non-LTR human families of transposable elements that have a mean copy number of >400,000, compared with 50, or less, for the small genomes. The mean family copy number of both the LTR and DNA elements in humans is >4000 per family which is at least one or two orders of magnitude higher than that for almost all the transposable element families of small genomes. Overall, the mean copy number per family of human transposable elements is 12,000, that is, between 2 to 3 orders of magnitude higher than that for the small genomes.

Maize genome transposable element comparison

Although no large plant genome has yet been sequenced, extrapolation from a 225-kb sequence to the whole *Zea mays* (maize) genome (2500 Mbp) provides some approximate comparative information (SanMiguel et al., 1996; Tikhonov et al., 1999). Genes occupy approximately 20% of the sequenced region. In contrast, interspersed repetitive elements, mostly transposable elements, make up 50–75% of the region. The transposable elements observed primarily consist of LTR retroelements that are often nested inside one another in intergenic regions. The most abundant transposable elements are six LTR retroelement families present in high copy number. Perhaps surprisingly, these abundant elements found between genes are largely intact and simply organized. In addition to the predominant LTR retroelements, some DNA elements, represented by *MITE*s, occupy low copy number sequences close to genes. Interestingly, not a single *MITE* was identified within the space occupied by the LTR retroelements in the sequenced *Adh-1* region (Bennetzen et al., 1998).

It is thought that the sequenced *Adh-1* region is fairly typical of the maize genome as a whole (Bennetzen et al., 1998; Tikhonov et al., 1999). However, see Fu et al. (2001) for evidence that this genome includes at least one unusually gene-rich region. Extrapolation of the data from the *Adh-1* region to the whole genome leads to copy numbers ranging from 2000 to 30,000 per family for the most abundant LTR retrotransposons (SanMiguel et al., 1996; Bennetzen et al., 1998). Other families of LTR retroelements with lower copy numbers are estimated to account for at least 10% of the maize nuclear DNA (SanMiguel et al., 1996). The wide range in family size for maize transposable elements reported for the *Adh-1* region is confirmed by independent studies. For example, the LTR retrotransposon *PREM-2* was estimated to be present in more than 10,000 copies, occupying ∼5% of the maize genome (Kumar & Bennetzen, 1999). In contrast, the LTR retrotransposon, *Bs1*, is estimated to be present in approximately two genomic copies (Jin & Bennetzen, 1989).

In summary, the distribution of transposable elements in the maize genome is very different from that of humans. In maize, LTR retroelements overwhelmingly predominate. In humans, non-LTR retroelements (*LINE*s and *SINE*s) predominate, but both LTR and DNA elements are also well represented. The two genomes are similar, however, in that the copy numbers of the largest transposable element families are greater, by at least two orders of magnitude, than those for the five smaller genomes.

Intraspecific variation in genome size

Perhaps surprisingly, it appears that the size of some transposable element families can wax and wane quite rapidly within a single species. Consequently, the proportion of the genome occupied by transposable elements is not constant within species, or even within the

same local population. This has recently been dramatically illustrated in wild barley, *Hordeum spontaneum*, in which the proportion of the genome occupied by the *BARE-1* retrotransposon varies in different ecological regions of Evolution Canyon in Israel (Kalendar et al., 2000). Full-length elements showed a 3-fold range of copy number (8.3 to 22.1×10^3 per haploid genome equivalent) among individuals at a single microsite. The data support an association between *BARE-1* copy number and the ecogeography of the canyon. Copy number is higher in the upper, drier sites of the canyon, that present greater stress to the plants, than the lower slopes. The implication is that the proliferation of the *BARE-1* element contributes to genome size evolution within and among local populations. It is unknown at this time whether the associations described are correlative or causal, and whether *BARE-1* copy number variation represents a direct, or indirect, response to environmental variation.

Extensive earlier studies of genome size in mosquitoes have demonstrated a large degree of intraspecific variation. Analysis of 47 geographical populations of *A. albopictus* from 18 countries showed a 2.5-fold variation in haploid genome size, ranging from 620 to 1660 Mbp (Rai & Black, 1999). It was claimed that genome size differences between two *Ae. albopictus* strains were due to differences in the amounts of highly repetitive DNA (Black & Rai, 1988), but further comparative studies are needed to differentiate among the various types of repetitive sequences present, using modern genome sequencing techniques. Intraspecific variation in total genome size has also been reported in a number of angiosperm species (Bennett & Smith, 1976).

Unassigned genome sequences

Many transposable elements have been shown to be ancient components of eukaryotic genomes. Because many of them undergo a degenerative process over evolutionary time, it is expected that their remnants will eventually diverge to the point of being unrecognizable as transposable elements. Interspersed repeats degraded to the point of being unrecognizable might therefore make up a significant fraction of unassigned genome sequences (Henikoff et al., 1997). In many *Drosophila* species, the centric heterochromatin has often been considered to be a graveyard for dead transposable elements because of the low frequency of recombination in this region.

Indirect contributions of transposable elements to interspecific variation in genome size

Quantitative measurements of the direct contributions of transposable elements to genome size, as described in the last section, are likely to underestimate the total contribution made by these elements because of their many indirect effects that are not measurable. A few examples are provided below, first in terms of mechanisms that may affect any part of the genome, and second, in terms of specific subgenomic fractions, which may be influenced by the presence of transposable elements.

Global genomic mechanisms

Ectopic recombination

There is growing evidence that transposable elements play an important role in refashioning genomic architecture. In addition to ectopic or homologous recombination (Caceres et al., 2001), alternative transposition of Class II elements may also lead to chromosomal rearrangements (Gray, 2000). Ectopic recombination between copies of the same transposable element family, located in nonhomologous genomic locations, can lead to duplications, deletions and new linkage relationships, often with consequent fitness reduction. This can potentially lead to selection against increased copy number and is considered by some (e.g., Charlesworth et al., 1997) to be a more important factor in containing transposable element copy number than selection against insertion mutations (Biémont et al., 1997). The theory predicts that because of their propensity for ectopic recombination, transposable elements will be found more frequently in chromosomal regions having lower recombination, such as in centromeric and telomeric heterochromatin, than in other chromosomal regions (Langley et al., 1988).

A recently reported example illustrates that ectopic recombination can lead not only to chromosomal rearrangements, such as inversions (Caceres et al., 1999), but also to a surprising degree of local genomic instability. Massive restructuring was found to be associated with the breakpoints of a transposable element-induced inversion in *D. buzzatii* (Caceres et al., 2001).

Transduction

Some transposable elements can provide a vehicle for the mobilization of flanking nucleotide sequences accompanying aberrant transposition events. In addition

to non-genic sequences, gene sequences such as those of exons or promoters, can sometimes be transduced and inserted into other existing genes. This may provide a general mechanism for the evolution of new genes. An example of so-called 'exon retroshuffling', that is, exon shuffling, mediated by the human *L1* element, was reported in tissue culture cells (Moran et al., 1999). More recently, two instances of 5′ transduction were reported, involving sequences 145 and 215 bp in length, following the sequencing of the human genome (International Human Genome Sequencing Consortium, 2001). This provided confirmation that active *L1* elements can transduce host DNA sequences and move these sequences to new genomic locations. Other Class I elements and some Class II elements in both animals and plants were earlier shown to acquire or transduce cellular genes (e.g., Bureau et al., 1994; Boeke & Stoye, 1997; Yu et al., 2000). Quantitative estimates of the frequency of transduction suggest that this process is responsible for the origin of 0.6–1% of human genome sequences (Goodier et al., 2000; Pickeral et al., 2000).

Gene duplication

Strong evidence is continuing to accumulate to support the proposal of Ohno (1970) that gene duplication has been of major importance in evolution. However, despite the myriads of cases of gene duplication identified, the mechanisms involved are not well understood. Rather than the predominant tandem repeats hypothesized by Ohno (1970) as being the source of new genes, analysis of the human genome indicates that dispersed gene duplications are considerably more common (Green & Chakravarti, 2001). Retroposition has been implicated as the most common mechanism for the creation of new genes in the human genome (Green & Chakravarti, 2001). When this occurs, processed mRNAs are retrotransposed and intron-less paralogs are inserted into new genomic locations. *LINE*s and LTR retrotransposons have been suggested as the most common source of reverse transcriptase required for catalyzing this process. A number of segmental duplications have also been identified in the human genome, in which entire genomic segments are duplicated and translocated to a new site.

Indirect subgenomic consequences of transposable element activity

Here are described some of the ways that transposable elements may contribute to genome size, but in an indirect way through their effects on various genomic constituents. In most cases it is impossible to provide quantitative estimates of these effects.

Microsatellites

Microsatellites are hypervariable DNA sequences made up of tandemly repeated short motifs. An association between transposable elements and microsatellites has been most frequently documented for various families of *SINE*s in mammals. For example, it is thought that retrotranscripts of *Alu* elements in humans undergo 3′ polyadenylation prior to their incorporation into the genome. This gives rise to a strong association between polyA microsatellites and the 3′ ends of these *SINE*s. This may serve as a guide to the retroposition of these elements into the genome (Nadir et al., 1996). In barley Ramsay (1999) found a close association of microsatellite repeats with retrotransposons and other dispersed repetitive elements.

Wilder and Hollacher (2001) recently identified a novel class of Dipteran mobile elements, called *mini-me* elements. These elements contain two internal proto-microsatellite regions, one of which commonly expands into lengthy microsatellite repeats. *Mini-me* elements are highly abundant in many *Drosophila* species, accounting for approximately 1.2% of the *D. melanogaster* genome. Their high copy number gives them the potential to be a prolific source of microsatellite DNA variation and, indirectly, a source of variation in genome size. It was determined (Wilder & Hollocher, 2001) that microsatellites are generated within *mini-me* elements through two separate mutational processes, the expansion of preexisting tandem repeats and the conversion of sequences with high cryptic simplicity into tandemly repetitive DNA.

Heterochromatin and satellite DNAs

Heterochromatin is well represented in eukaryotic genomes. For example, as much as 30% of the *D. melanogaster* genome is made up of heterochromatin, as is 15% of the human genome (John, 1988). Heterochromatin is rich in satellite DNA and transposable elements as documented for species as diverse as *D. melanogaster* (Pimpinelli et al., 1995), *Arabidopsis* (Copenhaver & Preuss, 1999), and maize (Ananiev et al., 1998).

A number of examples of ways in which transposable elements have contributed to the formation of heterochromatin have been described. For example, it has been demonstrated that gene silencing can be induced

by transposable elements (Dorer & Henikoff, 1994; Fanti et al., 1998). Further, Steinemann and Steinemann (1998) presented compelling evidence that the first step in Y chromosome degeneration is driven by the accumulation of transposable elements, especially retrotransposons. They suggest that an enrichment of these elements along an evolving Y chromosome could account for the switch from a euchromatic to a heterochromatic chromatin structure. This kind of experiment has lead to the suggestion that the evolution of heterochromatin may have been a response to invasive transposable elements.

Heterochromatin proteins can recognize and silence transposable elements (Fanti et al., 1998) some of which are known to target heterochromatin for insertion (Tschiersch et al., 1994). Thus, the evolution of heterochromatin could have lead to a self-perpetuating expansion of domains rich in transposable elements.

The origin of many families of satellite DNA (satDNA) is not known, but a few have been derived from transposable elements or have a major component that is related to a part of a mobile element. For example, the *pvB370* satDNA family (Heikkinen et al., 1995) is located in the centromeric heterochromatin of a number of species of the *D. virilis* group. This sequence was shown to have significant similarity to the long, direct, terminal repeats of the *pDv* transposable element family (Evgen'ev et al., 1982; Zelentsova et al., 1986). Also, the main heterochromatic satDNA of Cetaceans is made up of *LINE*-like repeats (Kapitonov et al., 1998).

A particularly interesting example of the potential of some transposable elements to generate other types of repetitive DNA is provided by the *SGM-IS* elements in *D. subobscura*, *D. guanche*, and *D. maderiensis* (Miller et al., 2000). *SGM-IS* elements show the characteristics of recent mobility, and are dispersed throughout the euchromatin. They have some of the properties of *MITE*s, such as the absence of homology to RT and integrase motifs and the potential to form stable secondary structures. *SGM-IS* elements apparently gave rise to *SGM-sat* sequences to which they are closely related (Miller et al., 2000). *SGM-IS* sequences comprise a major satDNA of *D. guanche* and make up 10% of the genome of this species (Miller et al., 2000). Also, among several previously unknown *En/Spm*-like families of transposable elements that were revealed by an analysis of the genome of *A. thaliana* (Kapitonov & Jurka, 1999) was an *En/Spm*-like family (*Atenspm*) which was found to be involved in generating satellite arrays in paracentromeric regions.

Simple sequences

Alu elements appear to have arisen in the human genome within the last 65 myr (Deininger & Batzer, 1999), and represent more than 10% of the genome. Almost all of the recently integrated *Alu* elements in the human genome belong to one of four closely related subfamilies. Analysis of the middle A-rich region of different *Alu* Ya5 members indicates a tendency towards expansion of this region and subsequent generation of simple sequence repeats (Roy et al., 2000).

Telomeres

Two non-LTR retroelements, *HetA*, and *TART* elements transpose specifically to the ends of *D. melanogaster* chromosomes and are present in tandem arrays at the ends of these chromosomes in place of the more typical telomere repeat sequences (Levis et al., 1993). Maintenance of the ends of chromosomes following DNA replication is achieved by the serial addition of these elements. Other insects, such as *Bombyx mori* also have telomere-specific non-LTR retrotransposons (Okazaki et al., 1995).

Discussion

Although several authors have previously proposed that transposable elements might provide a major part of the explanation for the C-value paradox in organisms such as plants, it has not been clear whether this relationship might extend to all organisms. Also, the precise way in which transposable elements might account for the variation in genome sizes has received only little discussion. Here, despite the very limited data that are currently available, I have attempted to start to examine these and other questions. The aim has not been to obtain definitive results, but to point the way to meaningful questions and approaches.

The preliminary data presented do appear to confirm that transposable elements play an important role in the determination of genome size, at least for large genomes. It will be interesting to see whether a linear relationship still holds, between the fraction of transposable elements and genome size, when data for more and diverse genomes are available, or whether a logistic curve will provide a better fit. The available preliminary data also indicate that transposable element copy number increases proportionately more

than number of transposable element families when smaller genomes are compared with larger ones. Both the mean and variance of transposable element copy number increase, often dramatically, in larger genomes compared with smaller ones. Copy number is uniformly quite low (<100) for most transposable element families found in the five small genomes analyzed. However, in the human genome the copy numbers of some transposable element families are several orders of magnitude larger. The key difference may be that restraints on copy number expansion have been lifted sufficiently for many, but not all, of the families found in the large genomes. It may be that transposition is globally repressed throughout small genomes, in contrast to the significant expansion that has apparently been possible for many transposable element families in the human and maize genomes. Another possibility is that the accumulation of large numbers of retroelements in the genomes of some species may be related to the presence of efficient silencing mechanisms, such as methylation, in these genomes. Such mechanisms may reduce the selective cost of maintaining large numbers of elements (McDonald, 1998).

Despite the assertion that number of transposable element families, as well as increased copy numbers accompany genome size expansion (Kumar & Bennetzen, 1999), no convincing indication was found in the limited data set examined here that the number of transposable element families is a major factor in genome expansion. However, with the continuing discovery of new families, this generalization may not hold. Also, transposable element families may be defined in different ways in different organisms and it will be important to obtain a general consensus concerning how these families should be defined. No single category of transposable elements is observed to predominate when the limited number of large genomes is compared. For example, non-LTR retroelements are predominant in humans, but are a very minor feature in maize in which the LTR retroelements overwhelmingly predominate. However, because of their dependence on *LINE*s for transposition, in order for *SINE*s to expand in humans, active copies of their companion *LINE*s would also need to increase.

It will be interesting to see if DNA elements will be found to predominate in any large genomes to be analyzed in the future. In the human sequence it was found that DNA elements, in addition to their small physical size, tend to be short-lived, on an evolutionary timescale, relative to the retroelements. This observation was interpreted in terms of their transposition mechanism and evolutionary life style (International Human Genome Sequencing Consortium, 2001). DNA elements are expected to transfer horizontally to another species relatively frequently in order to avoid eventual extinction (Kidwell, 1993). According to this scenario, DNA transposons are not expected to survive long enough to contribute to a major genome size expansion over long periods of evolutionary time, but they should leave proportionately more bits of themselves behind as footprints, or remnants. The sequencing of a number of additional large genomes should provide a test of this hypothesis.

Wong et al. (2000) assert that animals and plants have different genomic patterns for insertion of repeats. (Although the use of the term 'repeat' in this case covers a variety of sequence types, it clearly includes transposable elements.) Specifically, they conclude that in animals, most repeats integrate into intron DNA, but in plants, most repeats integrate into intergenic DNA. With increasing numbers of sequenced genomes, it will be interesting to see if this generalization is valid and whether it holds equally for different types of repeats, including transposable elements.

In conclusion, although the major focus of this paper has been on quantitative estimates of the bulk contribution of transposable elements to genome size, these elements make many contributions to genome evolution that cannot be quantified (e.g., McDonald, 1998; Bennetzen, 2000; Kidwell & Lisch, 2001). These are well summarized by Smit (1999) as follows: 'Far from merely expanding genomes with interspersed repeats, their legacy ranges from spliceosomal introns and antigen-specific immunity to many recent recruits in highly specialized functions'. In addition, if genome size is adaptive under certain conditions (Gregory & Hebert, 1999), transposable elements may represent an important mechanism for providing some of the variation in this trait on which natural selection acts.

Acknowledgements

I thank Tom Bureau, Andrew Holyoake, Damon Lisch, Dmitri Petrov, Jake Tu and two anonymous reviewers for helpful suggestions on the manuscript. Peter O'Donald provided valuable advice on the statistical analyses. This work was supported by National Science Foundation grant DEB 9815754.

References

Adams, M.D., S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, R.A. George, S.E. Lewis, S. Richards, M. Ashburner, S.N. Henderson, G.G. Sutton, J.R. Wortman, M.D. Yandell, Q. Zhang, L.X. Chen, R.C. Brandon, Y.H. Rogers, R.G. Blazej, M. Champe, B.D. Pfeiffer, K.H. Wan, C. Doyle, E.G. Baxter, G. Helt, C.R. Nelson, G.L. Gabor, J.F. Abril, A. Agbayani, H.J. An, C. Andrews-Pfannkoch, D. Baldwin, R.M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E.M. Beasley, K.Y. Beeson, P.V. Benos, B.P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M.R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K.C. Burtis, D.A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J.M. Cherry, S. Cawley, C. Dahlke, L.B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A.D. Mays, I. Dew, S.M. Dietz, K. Dodson, L.E. Doup, M. Downes, S. Dugan-Rocha, B.C. Dunkov, P. Dunn, K.J. Durbin, C.C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A.E. Gabrielian, N.S. Garg, W.M. Gelbart, K. Glasser, A. Glodek, F. Gong, J.H. Gorrell, Z. Gu, P. Guan, M. Harris, N.L. Harris, D. Harvey, T.J. Heiman, J.R. Hernandez, J. Houck, D. Hostin, K.A. Houston, T.J. Howland, M.H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G.H. Karpen, Z. Ke, J.A. Kennison, K.A. Ketchum, B.E. Kimmel, C.D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A.A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T.C. McIntosh, M.P. McLeod, D. McPherson, G. Merkulov, N.V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S.M. Mount, M. Moy, B. Murphy, L. Murphy, D.M. Muzny, D.L. Nelson, D.R. Nelson, K.A. Nelson, K. Nixon, D.R. Nusskern, J.M. Pacleb, M. Palazzolo, G.S. Pittman, S. Pan, J. Pollard, V. Puri, M.G. Reese, K. Reinert, K. Remington, R.D. Saunders, F. Scheeler, H. Shen, B.C. Shue, I. Siden-Kiamos, M. Simpson, M.P. Skupski, T. Smith, E. Spier, A.C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A.H. Wang, X. Wang, Z.Y. Wang, D.A. Wassarman, G.M. Weinstock, J. Weissenbach, S.M. Williams Woodage, T.K.C. Worley, D. Wu, S. Yang, Q.A. Yao, J. Ye, R.F. Yeh, J.S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X.H. Zheng, F.N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H.O. Smith, R.A. Gibbs, E.W. Myers, G.M. Rubin & J.C. Venter. 2000. The genome sequence of *Drosophila melanogaster*. Science 287: 2185–2195.

Ananiev, E.V., R.L. Phillips & H.W. Rines, 1998. Complex structure of knob DNA on maize chromosome 9. Retrotransposon invasion into heterochromatin. Genetics 149: 2025–2037.

Bennett, M.D. & J.D. Smith, 1976. Nuclear DNA amounts in angiosperms. Phil. Trans. R. Soc. Lond. B 274: 227–274.

Bennetzen, J.L., 2000. Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. 42: 251–269.

Bennetzen, J.L., P. SanMiguel, M. Chen, A. Tikhonov, M. Francki & Z. Avramova, 1998. Grass genomes. Proc. Natl. Acad. Sci. USA 95: 1975–1978.

Biémont, C., A. Tsitrone, C. Vieira & C. Hoogland, 1997. Transposable element distribution in *Drosophila*. Genetics 147: 1997–1999.

Black, W.C. & K.S. Rai, 1988. Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization. Genet. Res. 51: 185–196.

Boeke, J.D. & J.P. Stoye, 1997. Retrotransposons, Endogenous Retroviruses and the Evolution of the Retroelements in Retroviruses, edited by J.M. Coffin, S.H. Hughes & H.E. Varmus. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bureau, T.E., S.E. White & S.R. Wessler, 1994. Transduction of a cellular gene by a plant retroelement. Cell 77: 479–480.

Caceres, M., M. Puig & A. Ruiz, 2001. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. Genome Res. 11: 1353–1364.

Caceres, M., J.M. Ranz, A. Barbadilla, M. Long & A. Ruiz, 1999. Generation of a widespread *Drosophila* inversion by a transposable element. Science 285: 415–418.

Capy, P., C. Bazin, D. Higuet & T. Langin, 1997. Dynamics and Evolution of Transposable Elements. Landes Bioscience, Austin TX.

Charlesworth, B., C.H. Langley & P.D. Sniegowski, 1997. Transposable element distributions in *Drosophila*. Genetics 147: 1993–1995.

Copenhaver, G.P. & D. Preuss, 1999. Centromeres in the genomic era: unraveling paradoxes. Curr. Opin. Plant Biol. 2: 104–108.

Cresse, A.D., S.H. Hulbert, W.E. Brown, J.R. Lucas & J.L. Bennetzen, 1995. *Mu1*-related transposable elements of maize preferentially insert into low copy number DNA. Genetics 140: 315–324.

Deininger, P.L. & M.A. Batzer, 1999. *Alu* repeats and human disease. Mol. Genet. Metab. 67: 183–193.

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

Duret, L., G. Marais & C. Biemont, 2000. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. Genetics 156: 1661–1669.

Elgar, G., M.S. Clark, S. Meek, S. Smith, S. Warner, Y.J. Edwards, N. Bouchireb, A. Cottage, G.S. Yeo, Y. Umrania, G. Williams & S. Brenner. 1999. Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning. Genome Res. 9: 960–971.

Evgen'ev, M.B., G.N. Yenikolopov, N.I. Peunova & Y.V. Ilyin, 1982. Transposition of mobile genetic elements in interspecific hybrids of *Drosophila*. Chromosoma 85: 375–386.

Fanti, L., D.R. Dorer, M. Berloco, S. Henikoff & S. Pimpinelli, 1998. Heterochromatin protein 1 binds transgene arrays. Chromosoma 107: 286–292.

Feschotte, C. & C. Mouches, 2000. Recent amplification of miniature inverted-repeat transposable elements in the vector mosquito *Culex pipiens*: characterization of the Mimo family. Gene 250: 109–116.

Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. Trends Genet. 5: 103–107.

Fu, H., W. Park, X. Yan, Z. Zheng, B. Shen & H.K. Dooner, 2001. The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. Proc. Natl. Acad. Sci. USA 98: 8903–8908.

Garber, K., I. Bilic, O. Pusch, J. Tohme, A. Bachmair, D. Schweizer & V. Jantsch, 1999. The *Tpv2* family of retrotransposons of *Phaseolus vulgaris*: structure, integration characteristics, and use for genotype classification. Plant Mol. Biol. 39: 797–807.

Glockner, G., K. Szafranski, T. Winckler, T. Dingermann, M.A. Quail, E. Cox, L. Eichinger, A.A. Noegel & A. Rosenthal. 2001. The complex repeats of *Dictyostelium discoideum*. Genome Res. 11: 585–594.

Goodier, J.L., E.M. Ostertag & H.H. Kazazian Jr., 2000. Transduction of 3′-flanking sequences is common in *L1* retrotransposition. Hum. Mol. Genet. 9: 653–657.

Gray, Y.H., 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet. 16: 461–468.

Green, E.D. & A. Chakravarti, 2001. The human genome sequence expedition: views from the 'base camp'. Genome Res. 11: 645–651.

Gregory, T.R., 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol. Rev. Camb. Philos. Soc. 76: 65–101.

Gregory, T.R. & P.D. Hebert, 1999. The modulation of DNA content: proximate causes and ultimate consequences. Genome Res. 9: 317–324.

Heikkinen, E., V. Launonen, E. Muller & L. Bachmann, 1995. The *pvB370 Bam*III satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic *pDv* elements. J. Mol. Evol. 41: 604–614.

Henikoff, S., E.A. Greene, S. Pietrokovski, P. Bork, T.K. Attwood & L. Hood, 1997. Gene families: the taxonomy of protein paralogs and chimeras. Science 278: 609–614.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Jin, Y.K. & J.L. Bennetzen, 1989. Structure and coding properties of *Bs1*, a maize retrovirus-like transposon. Proc. Natl. Acad. Sci. USA 86: 6235–6239.

John, B., 1988. The biology of heterochromatin, pp. 1–128 in Heterochromatin, Molecular and Structural Aspects, edited by R.S. Verma. Cambridge University Press, Cambridge, UK.

Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo & A.H. Schulman, 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. USA 97: 6603–6607.

Kapitonov, V.V., G.P. Holmquist & J. Jurka, 1998. L1 repeat is a basic unit of heterochromatin satellites in Cetaceans. Mol. Biol. Evol. 15: 611–612.

Kapitonov, V.V. & J. Jurka, 1999. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica 107: 27–37.

Kapitonov, V.V. & J. Jurka, 2001. Rolling-circle transposons in eukaryotes. Proc. Natl. Acad. Sci. USA 98: 8714–8719.

Kidwell, M.G., 1993. Lateral transfer in natural populations of eukaryotes. Ann. Rev. Genet. 27: 235–256.

Kidwell, M.G. & D.R. Lisch, 2000. Transposable elements and host genome evolution. Trends Ecol. Evol. 15: 95–99.

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

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

Kumar, A., 1996. The adventures of the *Ty1-copia* group of retrotransposons in plants. Trends Genet. 12: 41–43.

Kumar, A. & J.L. Bennetzen, 1999. Plant retrotransposons. Annu. Rev. Genet. 33: 479–532.

Langley, C.H., E. Montgomery, R. Hudson, N. Kaplan & B. Charlesworth, 1988. On the role of unequal exchange on the containment of transposable element copy number. Genet. Res. 52: 223–235.

Le, Q.H., S. Wright, Z. Yu & T. Bureau, 2000. Transposon diversity in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 97: 7376–7381.

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

Lim, J.K. & M.J. Simmons, 1994. Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. Bioessays 16: 269–275.

Malik, H.S., S. Henikoff & T.H. Eickbush, 2000. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. 10: 1307–1318.

McClure, M.A., 1999. The retroid agents: disease, function, and evolution, pp. 163–195 in Origin and Evolution of Viruses, edited by E. Domingo, R. Webster & J. Holland. Academic Press, London.

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

Miller, W.J., A. Nagel, J. Bachmann & L. Bachmann, 2000. Evolutionary dynamics of the *SGM* transposon family in the *Drosophila obscura* species group. Mol. Biol. Evol. 17: 1597–1609.

Moran, J.V., R.J. DeBerardinis & H.H. Kazazian Jr., 1999. Exon shuffling by *L1* retrotransposition. Science 283: 1530–1534.

Nadir, E., H. Margalit, T. Gallily & S.A. Ben-Sasson, 1996. Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc. Natl. Acad. Sci. USA 93: 6470–6475.

Ohno, S., 1970. Gene Duplication. Springer Verlag, Berlin.

Okazaki, S., H. Ishikawa & H. Fujiwara, 1995. Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori*. Mol. Cell Biol. 15: 4545–4552.

Petrov, D.A., 2001. Evolution of genome size: new approaches to an old problem. Trends Genet. 17: 23–28.

Pickeral, O.K., W. Makaowski, M.S. Boguski & J.D. Boeke, 2000. Frequent human genomic DNA transduction driven by *LINE-1* retrotransposition. Genome Res. 10: 411–415.

Pimpinelli, S., M. Berloco, L. Fanti, P. Dimitri, S. Bonaccorsi, E. Marchetti, R. Caizzi, C. Caggese & M. Gatti. 1995. Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. Proc. Natl. Acad. Sci. USA 92: 3804–3808.

Rai, K.S. & W.C. Black, 1999. Mosquito genomes: structure, organization and evolution. Adv. Genet. 41: 1–33.

Ramsay, L., M. Macaulay, L. Cardle, M. Morgante, S. degli Ivanissevich, E. Maestri, W. Powell & R. Waugh, 1999. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J. 17: 415–425.

Roy, A.M., M.L. Carroll, S.V. Nguyen, A.H. Salem, M. Oldridge, A.O. Wilkie, M.A. Batzer & P.L. Deininger, 2000. Potential gene conversion and source genes for recently integrated *Alu* elements. Genome Res. 10: 1485–1495.

SanMiguel, P., B.S. Gaut, A. Tikhonov, Y. Nakajima & J.L. Bennetzen, 1998. The paleontology of intergene retrotransposons of maize. Nat. Genet. 20: 43–45.

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

Smit, A.F., 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. 9: 657–663.

Steinemann, M. & S. Steinemann, 1998. Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. Genetica 103: 409–420.

The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature 408: 796–815.

Tikhonov, A.P., P.J. SanMiguel, Y. Nakajima, N.M. Gorenstein, J.L. Bennetzen & Z. Avramova, 1999. Colinearity and its exceptions in orthologous adh regions of maize and sorghum. Proc. Natl. Acad. Sci. USA 96: 7409–7414.

Tilford, C.A., T. Kuroda-Kawaguchi, H. Skaletsky, S. Rozen, L.G. Brown, M. Rosenberg, J.D. McPherson, K. Wylie et al., 2001. A physical map of the human Y chromosome. Nature 409: 943–945.

Tschiersch, B., A. Hofmann, V. Krauss, R. Dorn, G. Korge & G. Reuter, 1994. The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)3–9* combines domains of antagonistic regulators of homeotic gene complexes. EMBO J. 13: 3822–3831.

Tu, Z., 1997. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. Proc. Natl. Acad. Sci. USA 94: 7475–7480.

Tu, Z., 2000. Molecular and evolutionary analysis of two divergent subfamilies of a novel miniature inverted repeat transposable element in the yellow fever mosquito, *Aedes aegypti.* Mol. Biol. Evol. 17: 1313–1325.

Tu, Z., 2001a. Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. Proc. Natl. Acad. Sci. USA 98: 1699–1704.

Tu, Z., 2001b. Maque, a family of extremely short interspersed repetitive elements: characterization, possible mechanism of transposition, and evolutionary implications. Gene 263: 247–253.

Turcotte, K., S. Srinivasan & T. Bureau, 2001. Survey of transposable elements from rice genomic sequences. Plant J. 25: 169–179.

Vicient, C.M., A. Suoniemi, K. Anamthawat-Jonsson, J. Tanskanen, A. Beharav, E. Nevo & A.H. Schulman, 1999. Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. Plant Cell 11: 1769–1784.

Vieira, C., D. Lepetit, S. Dumont & C. Biemont, 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. Mol. Biol. Evol. 16: 1251–1255.

Waterston, R. & J. Sulston, 1995. The genome of *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 92: 10836–10840.

Wendel, J.F. & S.R. Wessler, 2000. Retrotransposon-mediated genome evolution on a local ecological scale. Proc. Natl. Acad. Sci. USA 97: 6250–6252.

Wilder, J. & H. Hollocher, 2001. Mobile elements and the genesis of microsatellites in dipterans. Mol. Biol. Evol. 18: 384–392.

Wong, G.K., D.A. Passey, Y. Huang, Z. Yang & J. Yu, 2000. Is 'junk' DNA mostly intron DNA? Genome Res. 10: 1672–1678.

Yu, Z., S.I. Wright & T.E. Bureau, 2000. Mutator-like elements in *Arabidopsis thaliana*. Structure, diversity and evolution. Genetics 156: 2019–2031.

Zelentsova, E.S., R.P. Vashakidze, A.S. Kraev & M.B. Evgen'ev, 1986. Dispersed repeats in *Drosophila virilis*: elements mobilized by interspecific hybridization. Chromosoma 93: 469–476.

Zhang, Q., J. Arbuckle & S.R. Wessler, 2000. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize. Proc. Natl. Acad. Sci. USA 97: 1160–1165.