PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Population Biology of Transposable Elements

J. F. Y. Brookfield

Phil. Trans. R. Soc. Lond. B 1986 **312**, 217-226
doi: 10.1098/rstb.1986.0003

Phil. Trans. R. Soc. Lond. B **312**, 217–226 (1986)
Printed in Great Britain

The population biology of transposable elements

By J. F. Y. Brookfield

Department of Genetics, School of Biological Sciences, University of Leicester, Leicester LE1 7RH, U.K.

A transposable element can be defined as a DNA sequence capable of moving to new sites in the genome. Such DNA sequences have been described in a wide range of organisms. The evolutionary processes affecting transposable elements can thus be divided into two categories: changes in sequence and changes in genomic location. As with other types of evolutionary change, the nature of the evolutionary process will be reflected in the extent and type of genetic variation existing in wild populations. Quantitative models of the evolution of transposable element sequences and positions will be outlined, and related to relevant data.

The extent to which models designed to describe obvious transposable elements such as the mobile sequences of *Drosophila* are also applicable to interspersed repetitive DNAs from other species will be discussed.

Introduction

Population genetics has traditionally been based upon the assumption that genetic variability within populations consists entirely of allelic variation at genetic loci which behave in a Mendelian way. Even for polygenic characters, on which the individual effects of specific loci influencing quantitative traits cannot be identified, it has nevertheless been understood that the underlying genetic variation consists of simple alternative forms of genes differing by one or a few base pairs in their DNA sequences. The consequences of this assumption has been that mutation rates postulated in population genetics theory have been constant in space and time and, perhaps even more importantly, sufficiently small that phenotypic changes in the characteristics of populations of organisms are always brought about by natural selection, and not by mutation pressure. The empirical support for this view has been the observed low rates of spontaneous mutation to phenotypically distinct alleles at major gene loci in *Drosophila*, *Escherichia coli*, and other genetically well-characterized organisms. These observed low mutation rates have themselves been explicable in terms of the low rate of base substitution in non-mutagenized organisms.

Thus the conclusion of the irrelevance of the rate of mutation to phenotypically important (as opposed to selectively neutral) evolutionary change arises from a view that mutation occurs at the low rates associated with base substitutions. However, it is becoming increasingly clear that many characteristics of living organisms, and genome size is only the most obvious of these, must evolve through mechanisms other than the fixation of base substitution mutations. The importance of mutation pressure in such processes is generally not known, but may be significant. This is particularly true of the evolution of the distributions of mobile DNA sequences.

Mobile DNA sequences

Around 15% of the genome of a typical eukaryote consists of interspersed repetitive DNA sequences (Bouchard 1982). It is virtually unthinkable that such repeated sequences could have arisen from unrelated DNA sequences by independent and convergent processes of base substitution occurring at diverse sites in the genome. Rather, these sequences must, at some time, have duplicated and moved to their present locations, and their similarities must be due to their common ancestry. Among the best characterized interspersed repetitive DNAs are those of *Drosophila melanogaster* (Rubin 1983). These sequences have been shown to be mobile in the genome since they differ in their locations between laboratory strains (Young 1979; Strobel 1982). Many of the 50–100 families of *D. melanogaster* mobile sequences show the *copia*-like DNA structure, in that they have a sequence length between 4 and 9 kilobases, and long terminal repeats of 300–500 bases in the direct orientation (Rubin 1983). This structure is very similar to that of the integrated proviruses of vertebrate retroviruses. Furthermore, 'virus-like particles' from *Drosophila*-cultured cells have been shown to contain full-length *copia* RNAs (Shiba & Saigo 1983) and circular *copia* molecules, apparently generated by reverse transcription, have been isolated (Flavell 1984). Finally, for one *copia*-like element, *17.6*, DNA sequencing has revealed a retrovirus-like arrangement of open reading frames within the element, with the central reading frame encoding a protein homologous to a retroviral reverse transcriptase (Saigo *et al.* 1984). If *copia*-like sequences move to new genomic locations via a retrovirus-like mechanism including the reinsertion of the reverse transcripts of RNAs, two important consequences follow. First, there is the possibility that such sequences might be capable of movement between hosts in viral particles. No evidence for such horizontal transfer exists at present. Secondly, if movement between genomic sites occurs via an RNA intermediate, it is essentially certain that the transposition process is replicative. Thus transposition will be associated with an over-replication of mobile sequences relative to the rest of the genome.

This property of over-replication makes *Drosophila copia*-like sequences archetypal examples of the kinds of sequences discussed by Doolittle & Sapienza (1980), Sapienza & Doolittle (1981) and Orgel & Crick (1980) in their concept of 'selfish DNA'. They argued that sequences capable of over-replication could spread through genomes and populations without any natural selection in their favour. This is manifestly correct, and thus to understand in a quantitative way the evolution of these sequences there is a need for new types of population genetics theories, and attempt to produce such theories have begun (Langley *et al.* 1983; Charlesworth & Charlesworth 1983; Brookfield 1982*a*; Hickey 1982; Ohta 1981, 1983, 1984, 1985; Ohta & Kimura 1981; Slatkin 1985; Doolittle *et al.* 1984; Ginzberg *et al.* 1984). Such models share the feature that the rate of mutation (in the form of transposable element movement), and not merely the selection and drift processes acting on the products of mutation, is itself an important determinant of population genetic variation.

Of course, at any specific genetic locus the rate of mutation through transposable element insertion will be low. Even for the most mutable loci in hybrid dysgenesis it is less than 1% (Simmons *et al.* 1984; Engels 1983). This is less than the probable selective coefficients acting against the mutant alleles generated. Thus mutation pressure arising from transposable element insertions will not cause phenotypically deleterious allelic substitutions at individual loci. However, if one considers a different genetic quantity, that of the number of copies of a given transposable element family per genome, then this will have an upwards mutation rate through

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

transposition which will be comparable to the likely effects of natural selection acting upon this trait.

Parasitic DNAs

Sequences that can over-replicate relative to the rest of the genome will be capable of increasing their numbers even if they are selectively harmful. Indeed, if they are not to continue to increase their numbers in the genome there must be some compensating mechanism balancing their accumulation in transposition. Natural selection against individuals with many copies of a transposable element may act as such a mechanism. Charlesworth & Charlesworth (1983) have studied conditions under which a stable equilibrium can exist between replicative transposition and selection. They show that the important determinant of such an equilibrium is the relation between the fitness, w_n, of an individual with n copies of a transposable element, and n. Assuming a large variation among individuals in transposable element positions (as noted empirically for *copia*-like sequences by Montgomery & Langley (1983)), such that variation in copy number between individuals within populations is approximately Poisson, they show that an equilibrium between selection and transposition could exist when

$$\mathrm{d}(\ln w_n)/\mathrm{d}n = u - v,$$

where u is the transposition rate and v the rate of element deletion. An equilibrium value of n will thus be one where the above equation holds. For this to be a stable equilibrium, $\mathrm{d}^2(\ln w_n)/\mathrm{d}n^2 < 0$. This can be interpreted intuitively by saying that, for a stable equilibrium, fitness must decline more than multiplicatively with copy number. The implication of this is that if a transposable element arises in a species, it will increase its abundance in the genome until such a stable equilibrium copy number is reached. Since selection will be acting at this point, there will be a genetic load suffered by the host species.

The load will be approximately $1 - \mathrm{e}^{-(u-v)\bar{n}}$, where $\bar{n}$ is the mean element copy number at equilibrium. This genetic load implies that the elements are harmful, or, in other words, parasitic. Furthermore, while little genetic load is induced by a *copia*-like sequence family, with $n = 30$–50, and $(u-v)$ much less than 10^4 (as shown by the observed low rate of movement of these sequences), there would seem to be no reason *a priori* why there should not be other host species or elements for which the number of elements needed before equilibrium is attained is sufficiently large to impose a very great genetic load on the host species. Indeed, there may not be an equilibrium reached and transposable elements may increase without limit, reducing host fitness to zero.

If, however, interspersed repetitive DNAs are parasitic, why have hosts not themselves evolved mechanisms to eliminate them, thereby increasing their own fitness? One answer could be that the interspersed DNAs are well-adapted parasites and, as such, have been capable of overcoming any host attempt to delete them. However, such an explanation is clearly entirely circular, since our identification of interspersed repeats as well adapted is merely due to the observation that they do, in fact, persist. It cannot therefore, in itself, explain such persistence.

The problem, therefore, is that the view that organisms, parasitic or otherwise, are all well adapted fails to produce predictions of the natures or abundances of parasitic DNAs that compete with their hosts. This is a general and quite intractable problem (Brookfield 1982*b*). It is equally apparent in considerations of predator–prey or herbivore–plant interactions. A predator, for example, cannot be specifically well adapted to its prey species without those

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

features of its prey to which it is adapted thus being necessarily maladaptive to the prey. A naïve concept of pan-adaptationism is not merely unproductive in such a case; it is actually self-contradictory.

Thus considerations of the nature of parasitism cannot lead to empirical predictions of what proportion of the genome should consist of over-replicating DNA sequences and what structures they should show. It is possible, however, to take a more mechanistic view of such sequences and to make testable predictions of observable phenomena based upon assumptions about the dynamic properties of given sequences.

Variation in the genomic locations of transposable elements

What kinds of testable predictions can be made concerning transposable element distributions based upon dynamical models? One type of prediction has been of the frequency spectrum of transposable element sites in a finite population of hosts. In finite populations, genetic drift will act on each generation and genetic variability will be lost. This loss will be balanced by mutation, to produce an equilibrium level of variability. For transposable element position variation, the variation-generating process is transposition to new sites. Langley *et al.* (1983), and Charlesworth & Charlesworth (1983) have produced models of the population dynamics of transposable elements in a finite host population. They assume that elements transpose replicatively at a rate that varies negatively with copy number, that they are neutral with respect to natural selection and that they are also subject to a deletion process that precisely excises elements from the genome. They define this latter process as being at a copy-number-independent rate of μ per copy per generation. Thus when the copy number is such that the rate of transposition per element has been reduced to μ, there will be no expected change in copy number. Langley *et al.* (1983) further assume an effectively infinite number of occupiable genomic locations for transposable elements, so that repeated transpositions do not go to the same site, an assumption relaxed by Charlesworth & Charlesworth (1983).

Langley *et al.* (1983) show that, provided there is a low rate of immigration into the population, so that elements do not stochastically become extinct, there will be an equilibrium established with a characteristic frequency spectrum of sites. They show that the expected number of sites in the genome which have frequencies in the range from x to $x+\delta x$ is

$$\Lambda\theta x^{-1}(1-x)^{\theta-1}\,\delta x,$$

where Λ is the expected number of transposable elements per haploid genome, and $\theta = 4N_e\mu$, where N_e is the effective population size of the host.

This is a result analogous to the familiar infinite alleles frequency distribution of single-locus population genetics theory (Kimura & Crow 1964), and which forms a special case in the more general model of Charlesworth & Charlesworth (1983). Computer simulations confirm the accuracy of this equation, which essentially predicts great variability between individuals in transposable element position with $\theta \gg 1$, and very little variability with $\theta \ll 1$.

Montgomery & Langley (1983) have carried out *in situ* hybridization experiments with the transposable elements *297*, *copia*, and *412* as probes to 20 X chromosomes from a wild population of *Drosophila melanogaster*. Their data set is shown in table 1. There is very great variation between the X chromosomes studied in the position of the transposable elements, and θ estimates (produced by using the statistical techniques of Kaplan & Brookfield (1983)) were

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

16.7 for *297*, 35.0 for *412*, and 48.3 for *copia*. A problem, however, exists with data of this kind. *In situ* hybridization does not permit the exact localization of homologous sequences on the chromosome. An observed site could be anywhere within a region of 50–100 kilobases of DNA. Thus, possibly, transposable element sites apparently shared by two or more chromosomes in the sample could in reality be different sites sufficiently close together to be indistinguishable by the experimental technique. A calculation of how frequently such mis-scorings of sites might arise can be based upon a value for the 'effective number of distinguishable sites' on the X chromosome of around 100. The postulated mis-scoring events would be expected to occur sufficiently frequently to account for all the observations of sites shared between chromosomes (Kaplan & Brookfield 1983). The θ values given above thus considerably underestimate the true values. *Drosophila melanogaster* X chromosomes are thus extremely variable in their transposable element positions.

TABLE 1. DISTRIBUTION OF THREE TRANSPOSABLE ELEMENTS IN A SAMPLE OF X CHROMOSOMES SAMPLED FROM A WILD POPULATION OF *DROSOPHILA MELANOGASTER*

	sites occupied in a sample of 20 chromosomes				
element	once	twice	three times	four times	five times
copia	22	5	0	0	0
297	35	11	2	2	1
412	33	4	3	0	0

To what extent are *D. melanogaster* transposable elements typical of interspersed repetitive DNA sequences in eukaryotes? Most other interspersed repetitive sequences are shorter than those in *D. melanogaster*. In the human genome the most abundant interspersed repetitive sequence is the *Alu* sequence (Jelinek & Schmid 1982), which is approximately 290 base pairs in length and found around 300 000 times per haploid genome. Despite evidence for the mobility of this sequence (Jagadeeswaren *et al.* 1981; Hess *et al.* 1983) copies appear to be constant in position in population samples. For example, there are eight copies of the *Alu* sequence in the normal human β-like globin gene cluster (Allan & Paul 1983) but, in each of at least 250 haplotypes for the cluster examined to date all these copies but no others have been found (Jeffreys 1979; Antonorakis *et al.* 1984). The *Alu* sequence, therefore, does not provide a means of testing any prediction of the frequency spectrum except by noting that its invariance corresponds to a θ value very much less than 1.

A sequence intermediate in mobility between *Alu* and *copia* appears to be the interspersed repetitive sequence of *D. heteroneura* of less than 2.2 kilobases which has been cloned on plasmid pDh 3.5M (Hunt *et al.* 1984). This sequence shows a minority of sites of *in situ* hybridization shared between species of Hawaiian *Drosophila*, which indicates less positional variation than between *D. melanogaster* individuals for *copia*-like sequences.

DNA SEQUENCE EVOLUTION IN TRANSPOSABLE ELEMENT FAMILIES

The presence of homologous DNA sequences scattered throughout the genomes of higher organisms as interspersed repetitive DNAs implies that these sequences are, or have been in the past, capable of moving to new genomic locations. Transposition has thereby had the effect of creating scattered DNA sequences of shared descent. The predictions for the frequency

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

spectrum of transposable element sites given above are based upon an explicit model for the dynamics of transposition. Such a transposition model should, in addition, be capable of forming the basis of a model for the relation between transposable elements at different sites. Ohta (1984, 1985) has produced models for the effect of unbiased gene conversion and duplicative transposition (which have quantitatively similar effects) on the expected identity coefficients of non-allelic transposable element copies. A further model has been produced by Slatkin (1985).

Brookfield (1985) has considered the analogous question of the expected time to a common ancestor for copies of a transposable element family from different genomic locations. A prediction of this time, $\overline{T}$, can be derived from the model of Langley *et al.* (1983) with the additional assumption that all copies of a transposable element family are functionally equivalent in their transposition and deletion probabilities. The predicted value of $\overline{T}$ is given by

$$\overline{T} = \frac{\Lambda(1+\theta)}{2\mu}.$$

This is the expected time to a common ancestor for transposable elements from different genomic locations. Thus, if transposable element sequence evolution is largely neutral, and occurs at a rate of V nucleotide changes per nucleotide per generation, one would expect two randomly chosen copies of the transposable element family to differ by a proportion $2\overline{T}V$ of their bases. This figure, will, of course, only be accurate if it is considerably less than 1.

This can be illustrated by the consideration of special cases. The *Drosophila melanogaster copia*-like sequences yield high θ estimates, and so

$$\overline{T} \approx \frac{\Lambda\theta}{2\mu} = 2N_e\Lambda.$$

As Λ is 20–50 for these sequences, and Kreitman (1983) estimates the *Drosophila melanogaster* effective population size as 3×10^6, $\overline{T}$ will be around 10^8 generations. In conjunction with known rates of *Drosophila* sequence evolution (Langley *et al.* 1983; Bodmer & Ashburner 1984) this time is broadly consistent with the reported 5% sequence divergence between *copia*-like sequence family members (Spradling & Rubin 1981).

For the *Alu* sequence in man, θ estimates will be low, thus $\overline{T} \approx \Lambda/2\mu$ and, since μ is unknown but very much less than the reciprocal of the value of N_e of at least 10^4, and, since $\Lambda = 300000$, $\overline{T} \gg 10^{10}$ generations.

This is an absurdly high estimate of the time to a common ancestor and indicates that *Alu* sequence must be homogenized by further processes in addition to the expected homogenization by drift arising through the independent transpositions of 300000 functionally equivalent sequence copies. Candidates for such processes include biased gene conversion (and some evidence for *Alu* sequence conversion is known (Hess *et al.* 1984)), the selective replacement of all *Alu*s by a new variant with a transpositional or conversional advantage, contractions and expansions in the size of the *Alu* sequence family, and a process in which only one or a few *Alu* sequences serve as templates for all the others (Ullu & Tschudi 1984; Leigh Brown 1984).

Understanding of the processes whereby transposable DNA sequences come to be homogeneous will come through the systematic statistical analysis of DNA sequences of elements from many species and many families, and through the biochemical characterization of the transposition process.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The ecology of transposable elements

In *Drosophila* there are probably between 50 and 100 separate transposable element families which are all capable of movement to new genomic locations. Such families, in some cases (see, for example, Karess & Rubin 1984) encode transacting functional products. Others may move using transpositional machinery encoded by other elements or by host genes. The elements coexist in the environment comprising the chromosomes and cellular constitutions of, most crucially, the germ cells of their host species. They can be viewed accurately, if trivially, as being analogous to species in a complex ecological community. An important question in understanding their biology, therefore, is the extent to which elements of different sequence families interact with each other, and the extent to which their properties are dependent on the host genotype. Much evidence has accumulated to suggest that sequences are remarkably functionally independent. The fact that the different *copia*-like sequence families are flanked by short direct repeats of their target site DNAs of lengths characteristic for a family implies independence of the transpositional machineries used by different sequences (Spradling & Rubin 1981). For some mutations generated by transposable element insertions, the suppressability of such a mutation by a specific mutation at another locus is dependent upon the transposable element inserted and not upon the locus that is mutated (Jackson 1984; Modolell *et al.* 1983). The two systems of hybrid dysgenesis, I–R and P–M, are caused by non-homologous transposable DNA sequences (Rubin *et al.* 1982; Bucheton *et al.* 1984) and appear to be quite independent in their effects (Kidwell 1979).

Nevertheless, it would seem probable that, since most transposable element families are long-established components of the *Drosophila melanogaster* genome (Dowsett 1983; Martin *et al.* 1983; Brookfield *et al.* 1984) they might have evolved so as to interact in various ways. There is evidence that the sequences *copia* (Rubin *et al.* 1982) and *gypsy* (*mgd4*) (Gerasimova *et al.* 1984) move at elevated rates during P–M dysgenesis, and, indeed, it appears that in some conditions there may be the simultaneous transposition and deletion of many transposable element families (Gerasimova *et al.* 1984). In our laboratory we have isolated a strain of *Drosophila melanogaster* with a P–M hybrid dysgenesis-generated unstable mutation at *singed* which can revert somatically to wild type, giving areas of wild-type tissue in a *singed* fly. The genetics of this destabilization of *singed* are not fully elucidated, but we have shown that the instability does not map to the X-chromosome, and thus is not due simply to a mutational change in the presumed P-element located at *singed*. We have yet to show, however, whether this unusual somatic destabilization process results from a mutated host function or through the effects of a P-factor or P-element located elsewhere in the genome. In general terms, it may be possible to isolate many mutations in the host that will interfere with the functioning of transposable elements.

Many of the mutations that have arisen spontaneously in *Drosophila melanogaster* laboratory stocks over the last 70 years, and that have formed the basis of the study of *Drosophila* genetics, have been shown to be due to the insertion of transposable elements into or near structural genes. One important question concerns the extent to which such transposable element insertion mutations are used in adaptive evolution. If many spontaneous mutations are transposable element insertions, one would expect that occasionally such mutations may prove selectively advantageous, and, as a result, become fixed in the population by natural selection. Consequently, homology to the transposable element at a particular chromosomal site will come

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

to be shared by all individuals in the population. Thus, it is paradoxical that *in situ* hybridization experiments have failed to demonstrate any site occupied by the same transposable element insertion in all members of the *D. melanogaster* species. Why should it be that transposable element insertion mutations are not used in adaptive evolution? It cannot be that there is some feature shared by the phenotypes of transposable element insertion mutations which, in itself, makes such mutations invariably harmful, since Mackay (1984) has shown that genetic variability for polygenic characters, and not merely mutations of major effect, can be created by transposable element insertions. The phenotypic consequences of insertions will depend upon their locations relative to structural genes, and some must be potentially advantageous to an evolving species.

The resolution of this dilemma may lie, in part, in an imprecise excision process whereby transposable elements may be deleted along with part of the sequences flanking the insertion site, thus removing the transposable element sequence homology but not an advantageous mutant phenotype. Alternatively, it may be that the lack of fixed transposable element insertion sites truly reflects a low rate of adaptive evolution by the fixation of transposable element insertion mutations. It may indeed be that the overall rate of adaptive mutant substitutions per genome, which is a fundamental but quite unknown parameter of the evolutionary process, is itself as low as this supposed rate of adaptation through insertion mutation would suggest.

I thank C. H. Langley, B. Charlesworth and E. Montgomery for many useful discussions, M. Slatkin and T. Ohta for showing me unpublished manuscripts, and C. P. Kyriacou, G. P. Wallis, A. J. Jeffreys and T. Burke for useful comments on the manuscript.

References

Allan, M. & Paul, J. 1984 Transcription *in vivo* of the *Alu* family member upstream from the human ε-globin gene. *Nucl. Acids Res.* **12**, 1193–1200.

Antonorakis, S., Boehm, C. D., Sargeants, G. R., Theisen, C. E., Dover, C. J. & Kazazian, H. Jr 1984 Origin of the β^5 globin gene in blacks: the contribution of recurrent mutation or gene conversion or both. *Proc. natn. Acad Sci. U.S.A.* **81**, 853–856.

Bodmer, M. & Ashburner, M. 1984 Conversion and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. *Nature, Lond.* **309**, 425–430.

Bouchard, R. A. 1982 Moderately repetitive DNA in evolution. *Int. Rev. Cytol.* **76**, 113–193.

Brookfield, J. F. Y. 1982*a* Interspersed repetitive DNA sequences are unlikely to be parasitic. *J. theor. Biol.* **94**, 281–299.

Brookfield, J. F. Y. 1982*b* Adaptation and functional explanation in biology. *Evol. Theor.* **5**, 281–290.

Brookfield, J. F. Y. 1985 A model for DNA sequence evolution within transposable element families. (In the press.)

Brookfield, J. F. Y., Montgomery, E. & Langley, C. H. 1984 Apparent absence of transposable elements related to the P elements of *D. melanogaster* in other species of *Drosophila*. *Nature, Lond.* **310**, 330–332.

Bucheton, A., Paro, R., Sang, H. M., Pelisson, A. & Finnegan, D. J. 1984 The molecular basis of I–R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor. *Cell* **38**, 153–163.

Charlesworth, B. & Charlesworth, D. 1983 The population dynamics of transposable elements. *Genet. Res.* **42**, 1–27.

Doolittle, W. F., Kirkwood, T. B. L. & Dempster, M. A. H. 1984 Selfish DNAs with self-restraint. *Nature, Lond.* **307**, 501–502.

Doolittle, W. F. & Sapienza, C. 1980 Selfish genes, the phenotype paradigm and genome evolution. *Nature, Lond.* **284**, 601–603.

Dowsett, A. P. 1983 Closely related species of *Drosophila* can contain different libraries of middle repetitive DNA sequences. *Chromosoma* **88**, 104–108.

Engels, W. R. 1983 The P family of transposable elements in *Drosophila*. *A. Rev. Genet.* **17**, 315–344.

Flavell, A. J. 1984 Role of reverse transcription on the generation of extrachromosomal *copia* mobile genetic elements. *Nature, Lond.* **341**, 514–521.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Gerasimova, T. I., Ilyin, Y. V., Mizrokhi, L. J., Semjonova, L. V. & Georgiev, G. P. 1984 Mobilization of the transposable element *mdg4* by hybrid dysgenesis generates a family of unstable *cut* mutations in *Drosophila melanogaster*. *Molec. Gen. Genet.* **193**, 488–492.

Gerasimova, T. I., Mizrokhi, L. J. & Georgiev, G. P. 1984 Transposition bursts in genetically unstable *Drosophila melanogaster*. *Nature, Lond.* **309**, 714–716.

Ginzberg, L. R., Bingham, P. M. & Yoo, S. 1984 On the theory of speciation induced by transposable elements. *Genetics* **107**, 331–341.

Hess, J. R., Fox, M., Schmid, C. & Shen, C.-K. J. 1983 Molecular evolution of the human α-globin-like gene region: insertion and deletion of *Alu* family repeats and non-*Alu* DNA sequences. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5970–5974.

Hess, J. F., Schmid, C. W. & Shen, C.-K. J. 1984 A gradient of sequence divergence in the human α-globin duplication units. *Science, Wash.* **226**, 67–70.

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

Hunt, J. A., Bishop, J. G., III & Carson, H. 1984 Chromosomal mapping of a middle-repetitive DNA sequence in a cluster of five species of Hawaiian *Drosophila*. *Proc. natn. Acad. Sci. U.S.A*, **81**, 7146–7150.

Jackson, I. 1984 Transposable elements and suppressor genes. *Nature, Lond.* **309**, 751–752.

Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. 1981 Short interspersed repetitive DNA elements in eukaryotes: transposable DNA elements generated by reverse transcription of RNA Pol.III transcripts? *Cell* **26**, 141–142.

Jeffreys, A. J. 1979 DNA sequence variants in the G_γ-, A_γ-, δ- and β-globin genes of man. *Cell* **18**, 1–10.

Jelinek, W. R. & Schmid, C. W. 1982 Repetitive sequences in eukaryotic DNA and their expression. *A. Rev. Biochem.* **51**, 813–844.

Kaplan, N. L. & Brookfield, J. F. Y. 1983 Transposable elements in Mendelian populations. III. Statistical results. *Genetics* **104**, 485–495.

Karess, R. E. & Rubin, G. M. 1984 Analysis of P transposable element functions in *Drosophila*. *Cell* **38**, 135–146.

Kidwell, M. G. 1979 Hybrid dysgenesis in *Drosophila melanogaster*: the relationship between the P–M and I–R interaction systems. *Genet. Res.* **33**, 205–217.

Kimura, M. & Crow, J. F. 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Kreitman, M. 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature, Lond.* **304**, 412–416.

Langley, C. H., Brookfield, J. F. Y. & Kaplan, N. L. 1983 Transposable elements in Mendelian populations. I. A theory. *Genetics* **104**, 457–472.

Langley, C. H., Montgomery, E. & Quattlebaum, W. R. 1982 Restriction map variation in the *Adh* region of *Drosophila*. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5631–5635.

Leigh Brown, A. 1984 On the origin of the *Alu* family of repeated sequences. *Nature, Lond.* **312**, 106.

Mackay, T. F. C. 1984 Jumping genes meet abdominal bristles: hybrid dysgenesis-induced quantitative variation in *Drosophila melanogaster*. *Genet. Res.* **44**, 231–237.

Martin, G., Wiernasz, D. & Schedl, P. 1983 Evolution of *Drosophila* repetitive-dispersed DNA. *J. molec. Evol.* **19**, 203–213.

Modolell, J., Bender, W. & Meselson, M. 1983 *Drosophila melanogaster* mutations suppressible by the suppressor of hairy wing are insertions of a 7.3 kilobase mobile element. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1678–1682.

Montgomery, E. A. & Langley, C. H. 1983 Transposable elements in Mendelian populations. II. Distribution of three *copia*-like elements in a natural population of *Drosophila melanogaster*. *Genetics* **104**, 473–483.

Ohta, T. 1981 Population genetics of selfish DNA. *Nature, Lond.* **292**, 648–649.

Ohta, T. 1983 Theoretical study on the accumulation of selfish DNA. *Genet. Res.* **41**, 1–15.

Ohta, T. 1984 Population genetics of transposable elements. *J. Math. appl. Med. Biol.* **1**, 17–29.

Ohta, T. 1985 A model of duplicative transposition and gene conversion of repetitive DNA families. *Genetics* (In the press.)

Ohta, T. & Kimura, M. 1981 Some calculations on the amount of selfish DNA. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1129–1132.

Orgel, L. E. & Crick, F. H. C. 1980 Selfish DNA: the ultimate parasite. *Nature, Lond.* **284**, 604–606.

Rubin, G. M. 1983 Dispersed repetitive DNAs in *Drosophila*. In *Mobile genetic elements* (ed. J. A. Shapiro, pp. 329–361). New York: Academic Press.

Rubin, G. M., Kidwell, M. G. & Bingham, P. M. 1982 The molecular basis of P–M hybrid dysgenesis: the nature of the induced mutations. *Cell* **29**, 987–994.

Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. & Yuki, S. 1984 Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. *Nature, Lond.* **312**, 659–661.

Sapienza, C. & Doolittle, W. F. 1981 Genes are things you have whether you want them or not. *Cold Spring Harbor Symp. quant. Biol.* **45**, 619–628.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Shiba, T. & Saigo, K. 1983 Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosophila melanogaster*. *Nature*, *Lond.* **302**, 119–124.

Simmons, M. J., Raymond, J. D., Johnson, N. A. & Fahey, T. M. 1984 A comparison of mutation rates for specific loci and chromosome regions in dysgenic hybrid males of *Drosophila melanogaster*. *Genetics* **106**, 85–94.

Slatkin, M. 1985 Genetic differentiation of transposable elements under mutation and unbiased gene conversion. *Genetics* **110**, 145–158.

Spradling, A. C. & Rubin, G. M. 1981 *Drosophila* genome organisation: conserved and dynamic aspects. *A. Rev. Genet.* **15**, 219–264.

Strobel, E. 1982 Mobile dispersed repeated DNA elements in the *Drosophila* genome. *Fedn Proc. Fedn Am. Socs exp. Biol.* **41**, 42–44.

Ullu, E. & Tschudi, C. 1984 *Alu* sequences are processed 7SL RNA genes. *Nature*, *Lond.* **312**, 171–172.

Young, M. W. 1979 Middle repetitive DNA: a fluid component of the *Drosophila* genome. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6274–6278.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES