

## Transposable element copy number and stable polymorphisms

VIDYANAND NANJUNDIAH

Molecular Biology Unit, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

MS received 13 January 1986

**Abstract.** We consider a simple and analytically solvable model for the spread of a transposable element which has deleterious effects on fitness. Two possible modes are treated, one in which transposition occurs in the newly fertilized zygote, and another in which transposition takes place only in the germ line. In effect, transposition precedes selection in the first case and follows it in the second. This has different long-term consequences depending on the rate of transposition and the values of the selection coefficients. Conditions are derived for the existence of a stable polymorphism with respect to element copy number; the conditions are more stringent in the first case than in the second. It is proved that a polymorphism is impossible unless the copy number decreases fitness in a more-than-multiplicative fashion.

**Keywords.** Population genetics; evolution; transposons; polymorphism.

### 1. Introduction

A number of models relating to the dynamics of spread of transposable elements in a population have recently appeared in the literature (Brookfield 1982; Hickey 1982; Charlesworth and Charlesworth 1983; Langley *et al* 1983; Ginzburg *et al* 1984). The most interesting conclusion from these models is that in a sexually reproducing, or more precisely, outbreeding (Hickey 1982) population a transposable element can spread even if it decreases the fitness of an individual carrying it. Conditions have been derived giving the minimum frequency of transposition necessary to overcome adverse selection of a given intensity and ensure the spread of the element to eventual fixation. It has also been demonstrated that under certain conditions the population can exhibit a stable polymorphism with respect to the number of copies of the element carried by an individual (Langley *et al* 1983; Charlesworth and Charlesworth 1983). However, since the models which show this are either analytically complex or involve numerical simulation or both, it would be of help if the same results could be shown to hold in a simple scheme which can be treated analytically.

In this paper I work out the necessary and sufficient conditions for such a polymorphism to exist in the simplest possible situation, that of a transposon with a single site of insertion in a haploid genome. The basic model is due to Hickey (1982), who has drawn attention to, and made use of, its formal similarity to the standard one-locus-two-allele case of population genetics theory (e.g. Crow and Kimura 1970). It turns out that the conditions for the existence of a polymorphism differ depending on whether transposition occurs in the newly formed zygote (and therefore in all cells derived from it) or only in the germ line.

## 2. The model

We assume an infinitely large panmictic population. In the notation of Hickey (1982), let  $TT$ ,  $Tt$  and  $tt$  denote respectively individuals with two, one and zero copies of the transposable element. The symbol  $T$  can itself be conveniently taken to stand for the presence of the transposable element (inserted into "its" site),  $t$  for its absence. It is in this sense that  $T$  and  $t$  can also be imagined to be alternative alleles at a locus. In this model, transposition effectively converts a  $Tt$  individual into a  $TT$  one (it being understood that the transposable element is simultaneously duplicated, Harshey and Bukhari 1981). This can happen in one of two ways: either a newly fertilized  $Tt$  zygote can become  $TT$ , and both germ and somatic cells are  $TT$  [case *A*, the case treated by Hickey 1982], or, a primordial  $Tt$  germ cell changes to  $TT$ ; the somatic cells remain  $Tt$  while the germ-line cells are  $TT$  (case *B*). The means by which either of these processes might occur will not be discussed here.

We assume that there is a certain probability  $f$  ( $0 < f < 1$ ) per generation that transposition can occur, and that the relative fitnesses of  $TT$ ,  $Tt$  and  $tt$  individuals are  $1 - S_2$ ,  $1 - S_1$  and  $1$  respectively. Here  $0 < S_1 < S_2 < 1$  so that two copies of  $T$  lower fitness more than one copy of  $T$ . Let  $p$  be the population frequency of  $T$  to begin with; therefore that of  $t$  is  $1 - p$ . Tables 1 (case *A*) and 2 (case *B*) contain charts illustrating how transposition and selection affect the value of  $p$ . The essential difference between the two cases is that whereas transposition precedes selection in case *A*, it follows selection in case *B*. The reason being that in case *B* the somatic cells of the individual remain  $Tt$  in genotype, so the fitness of the individual is  $1 - S_1$ . "Effective genotype frequencies" in table 2 (case *B*) refers to the following fact: on account of transposition in the germ line a fraction  $f$  of  $Tt$  individuals are—with respect to hereditary transmission—in effect of genotype  $TT$ ; a fraction  $1 - f$ , in whom no transposition occurs, have the same actual and effective genotypes. In this indefinitely large population model,  $f$  can mean either the probability that all germ cells are changed from  $Tt$  to  $TT$  in an individual, or that a fraction  $x$  of germ cells is changed in  $f/x$  individuals. It should be noted that the meaning of  $f$  is different in cases *A* and *B*.

It now remains to calculate the changed values of  $p$  in the two cases and to examine in what manner  $p$  varies for different values of  $f$ ,  $S_1$  and  $S_2$ . Details of the calculations are sketched in the appendix. The method is to work out  $\Delta p$ , the change in the frequency of  $T$  in one generation, and equate it to zero. Suppose  $\Delta p = 0$  at  $p = p^*$ ,  $p^*$  lying between 0

**Table 1.** Relative genotype frequencies in the case of transposition in the one-celled zygote (case *A*).

Genotypes	$TT$	$Tt$	$tt$
Genotype frequencies initially	$p^2$	$2p(1-p)$	$(1-p)^2$
After transposition	$p^2 + f \cdot 2p(1-p)$	$(1-f) \cdot 2p(1-p)$	$(1-p)^2$
After selection	$p^2 + f \cdot 2p(1-p) \times (1-S_2)$	$(1-f) \cdot 2p(1-p) \times (1-S_1)$	$(1-p)^2$

In contrast to Hickey (1982), the selection coefficients ( $S_1$  and  $S_2$ ) are not frequency-dependent.

**Table 2.** Relative genotype frequencies when transposition occurs only in the germ line.

Genotypes	<i>TT</i>	<i>Tt</i>	<i>tt</i>
Genotype frequencies initially	$p^2$	$2p(1-p)$	$(1-p)^2$
After selection	$p^2(1-S_2)$	$2p(1-p)(1-S_1)$	$(1-p)^2$
Effective genotype frequencies after transposition	$p^2(1-S_2) + f \cdot 2p(1-p)(1-S_1)$	$(1-f) 2p(1-p) \times (1-S_1)$	$(1-p)^2$

Note that the transposition frequency,  $f$ , is measured per germ-line cell here whereas it is measured per zygote in case A (table 1).

and 1. If  $\Delta p > 0$  for  $p < p^*$  and  $\Delta p < 0$  for  $p > p^*$ , we conclude that  $p = p^*$  represents a stable polymorphism. If  $\Delta p$  is always positive (or  $p^* > 1$ ),  $T$  gets fixed in the population; if  $\Delta p$  is always negative (or  $p^* < 1$ )  $T$  gets eliminated.

### 3. Results

#### Case A: Transposition precedes selection

A stable polymorphism exists when

$$S_1/(1 + S_1 - 2S_2) < f < (S_2 - S_1)/(1 - S_1). \tag{1}$$

This is possible whenever  $S_2 < 0.5$  and  $S_1 < S_2(1 - 2S_2)/(2 - 3S_2)$ .

The equilibrium value of  $p$  is

$$p^* = \frac{1 - 2f(1 - S_2) - (1 - S_1)(1 - f)}{(1 - S_2)(1 - 2f) - 2(1 - S_1)(1 - f) + 1}. \tag{2}$$

Such a polymorphism is impossible in case the fitness is multiplicative, that is,  $(1 - S_2) = (1 - S_1)^2$ ; one needs  $(1 - S_2) < (1 - S_1)^2$ . In the special case where a single copy of the transposable element has no effect on fitness ( $S_1 = 0$ ) the various outcomes are easy to illustrate and are shown in figure 1.

#### Case B: Selection precedes transposition

The condition for a stable polymorphism to exist is

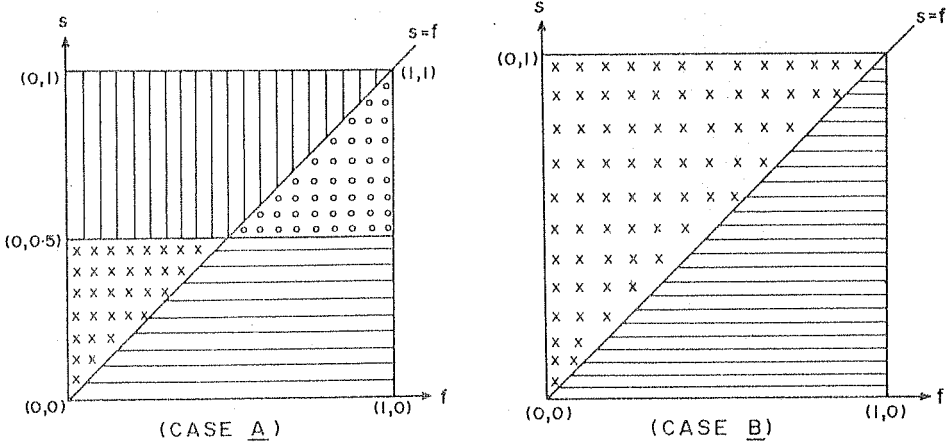
$$S_1/(1 - S_1) < f < (S_2 - S_1)/(1 - S_1), \tag{3}$$

a necessary condition for which is  $S_2 > 2S_1$ .

The equilibrium value of  $p$  is

$$p^* = [f - S_1(1 + f)]/(S_2 - 2S_1).$$

Once again, a stable polymorphism is impossible unless  $(1 - S_2) < (1 - S_1)^2$ . The results for the special case of  $S_1 = 0$  are illustrated in figure 2.



Figures 1 and 2. The various outcomes possible in the special case wherein one copy of the transposable element has no effect ( $S_1 = 0$ );  $S$  here stands for  $S_2$ . Transposition either occurs in the zygote (figure 1; case A) or in the germ line (figure 2; case B). The meaning of the symbols is as follows: horizontal lines, the transposable element gets fixed; vertical lines, gets eliminated; circles, gets fixed or eliminated depending on initial conditions; crosses, a stable polymorphism results. The selection coefficient is assumed constant; the relevant conditions under frequency-dependent selection can be derived as indicated in the appendix.

4. Discussion

In addition to the possibility that it might get eliminated or fixed, a transposable element which reduces individual fitness can persist in a population at an intermediate frequency. The consequence is a stable polymorphism, with the individuals in the population differing (in our model) according as they have no copies, one copy or two copies of the element. An examination of the conditions [(1) and (3)] shows that, loosely speaking, the polymorphism exists whenever the transposition frequency  $f$  is small relative to  $S_2$ , the intensity of selection against individuals carrying two copies of the element. Sufficiently high values of  $f$  lead to fixation. Very high values of  $S_2$  lead to elimination in case A; however, in case B, polymorphism persists even when the presence of two copies of the transposable element is lethal ( $S_2 = 1$ ). The interesting result, proven analytically, is that no polymorphism is possible when fitnesses are multiplicative; the fitness of a  $TT$  individual has to be less than the squared fitness of a  $Tt$  individual. This is in accordance with the finding of Charlesworth and Charlesworth (1983), using a mixture of analytical and numerical techniques, that "... either regulation of transposition rate in response to copy number per genome, a sufficiently strongly curved dependence of individual fitness on copy number, or both" would be needed to set up a polymorphism. In our model the question of regulation of transposition rate does not arise, of course, since transposition occurs only in the  $Tt$  state. The conditions for a polymorphism to exist are in some ways reminiscent of a mutation-selection balance, the difference being that here "mutations" can occur only in the heterozygote (or in its germ cells). Similarly, a reduction in copy number, caused by spontaneous or induced loss of the transposon, would be analogous to reverse mutation. For the sake of simplicity we have ignored the possibility of spontaneous loss; certainly at low values of  $p$  it cannot be important.

The two modes of transposition considered (cases *A* and *B*) differ in their consequences. In the first (case *A*, table 1, figure 1), what started out—at the time of fertilization—as a “heterozygote” for the transposable element becomes (with a certain probability) homozygous; if it does, the organism suffers a further loss in fitness. The second mode (case *B*, table 2, figure 2) is analogous to segregation distortion (Crow 1979) in that more than 50% of the gametes end up carrying two copies of the transposon; the effects on fitness are felt in the progeny. Allowing for differences in the meanings of the symbols used, the results derived here (case *B*) in fact follow directly from those of Hiraizumi *et al* (1960) in their treatment of meiotic drive in both sexes caused by a locus with properties similar to that of a *segregation distorter* in *Drosophila*. Table 3 displays equilibrium frequencies for two sets of values of  $S_1$ ,  $S_2$  and  $f$ ; clearly a range of distributions is possible.

The conditions for a polymorphism to exist are more stringent in case *A* than in case *B*. Indicative of this difference between the two cases is the fact that in case *B*, if the “heterozygote” suffers no loss in fitness ( $S_1 = 0$ ), the transposable element can never get eliminated (compare figures 1 and 2). An extreme consequence of this situation in case *B* is the following: if  $TT$  is a lethal condition ( $S_2 = 1$ ), the population gets polarised into two classes, a fraction  $2f/(1+f)$  surviving and containing one copy of  $T$ , and a fraction  $(1-f)/(1+f)$  containing none. Depending on  $f$ , the ratio of the numbers in the two classes can take any value from 0 to  $\infty$ ; for  $f = 0.33$ , the ratio is 1. Another sort of extreme polymorphism, illustrated in table 3 (last line), results in practically the entire population being made up of individuals carrying either one or two copies of the transposable element. These results are of interest in the light of the arguments put forward by Chandra (1985) that the essential difference between males and females is in the copy number of non-coding segments of DNA, (some of them) conceivably transposable elements.

In conclusion, it must be stressed that the model is intended to be merely illustrative of the general case of more than one kind of transposon and many sites of integration in the genome. It is well documented that in nature polymorphisms do exist with respect to copy numbers of transposable elements (Langley and Montgomery 1983).

Table 3. Illustrative chart showing the types of stable polymorphisms exhibited for two different sets of parameter values.

Parameter values	Case	Equilibrium frequency	Fraction of population in different classes		
			$TT$	$Tt$	$tt$
$S_1 = 0.01$ $S_2 = 0.04$ $f = 0.02$	<i>A</i>	0.46	0.21	0.50	0.29
	<i>B</i>	0.49	0.24	0.50	0.26
$S_1 = 0.08$ $S_2 = 0.40$ $f = 0.32$	<i>A</i>	0.27	0.07	0.39	0.53
	<i>B</i>	0.89	0.79	0.20	0.01

The values chosen for  $S_1$  and  $S_2$  are of the same order as those expected for deleterious mutations in *Drosophila melanogaster* (Simmons and Crow 1977).

It seems reasonable to generalise from the present results and infer that polymorphisms, such as those observed, will be reached if the several transposons can be treated in the manner of independently assorting alleles with more-than-multiplicative effects on fitness. In general, given a set of fitnesses and selection coefficients, more than one polymorphism should be possible, with the one actually reached determined by initial conditions. As a consequence, it might happen that the presence of (even) a single transposable element which has attained the maximal copy number consistent with its effect on fitness could act as a barrier to the invasion of the population by other elements. Thus, under natural conditions, potentially interbreeding sub-populations of the same species might come to differ to such an extent with respect to the classes of the transposable elements carried by them that they are in effect reproductively isolated from one another.

### Acknowledgement

I am grateful to Drs J F Crow and H Sharat Chandra for several helpful suggestions.

### Appendix

We begin from the expressions in the last rows of tables 1 and 2 and write down in each case the new frequency of  $T$ ,  $p^1$ . Therefore the change in frequency of  $p$  is given by  $\Delta p = p^1 - p$ . It turns out that  $\Delta p$  factors into a term which is always positive times a term whose sign—by implication—determines the existence and nature of the equilibrium. In what follows, we write  $q$  for  $1 - p$  where convenient.

Case A:

From table 1,

$$p^1 = \frac{(p^2 + f \cdot 2pq)(1 - S_2) + (1 - f)pq(1 - S_1)}{(p^2 + f \cdot 2pq)(1 - S_2) + (1 - f)2pq(1 - S_1) + q^2} \quad (\text{A1})$$

For the change in fitness this gives, after some algebra,

$$\begin{aligned} \Delta p &= p^1 - p \\ &= \frac{p(1 - p) [(p + f \cdot 2pq)(1 - S_2) + (1 - S_1)(1 - f)(1 - 2p) - (1 - p)]}{1 - 2pq [fS_2 + (1 - f)S_1]} \end{aligned} \quad (\text{A2})$$

Since  $f, S_1, S_2, p$  and  $q$  all lie between 0 and 1, the denominator in (A2) is  $> 0$ , as is the term  $p(1 - p)$ . The sign of  $\Delta p$  then depends on the expression

$$(p + 2fq)(1 - S_2) + (1 - S_1)(1 - f)(1 - 2p) - (1 - p). \quad (\text{A3})$$

At  $p = 0$ , this equals

$$f - 2fS_2 - S_1 + S_1f, \quad (\text{A4})$$

and at  $p = 1$ ,

$$1 - S_2 - (1 - f)(1 - S_1). \quad (\text{A5})$$

The requirement that (A4) be positive and (A5) negative gives condition (1) stated in the text. It follows directly that a subsidiary requirement is that

$$S_1 < (S_1)_{\max} = \frac{S_2(1 - 2S_2)}{(2 - 3S_2)}, S_2 < 0.5. \quad (\text{A6})$$

As an immediate consequence,

$$(1 - S_2) < [1 - (S_1)_{\max}]^2 \leq (1 - S_1)^2. \quad (\text{A7})$$

The value of  $p^*$  in (2) results from equating (A3) to zero. In the case of frequency-dependent selection  $S_1 = S_1(p)$  and  $S_2 = S_2(p)$ . The relevant conditions are obtained by using the appropriate values of  $S_1$  and  $S_2$  in (A4) and (A5). For instance, to take an extension of the case considered by Hickey (1982), if  $S_1 \equiv 0$  and  $S_2(0) < 0.5$ , the element will spread to fixation so long as  $f > S_2(1)$ ; in particular, if  $f = 1$ , it will do so even in the face of an adverse selection of practically 100%.

Case B

From table 2,

$$p^1 = \frac{p^2(1 - S_2) + f \cdot 2pq(1 - S_1) + (1 - f) \cdot pq \cdot (1 - S_1)}{p^2(1 - S_2) + 2pq(1 - S_1) + q^2}. \quad (\text{A8})$$

Therefore (after some simplification)

$$p = p^1 - p = \frac{p(1 - p) [-S_2p + f(1 - S_1) - S_1(1 - 2p)]}{1 - S_2p^2 - 2pqS_1}. \quad (\text{A9})$$

The denominator  $1 - S_2p^2 - 2pqS_1$  in (A9) is positive at  $p = 0$  as well as at  $p = 1$  (assuming that  $1 - S_2$  is positive); the expression being a quadratic in  $p$ , it follows that it can be negative at some intermediate value of  $p$  only if it has a negative minimum within this range. Elementary algebra shows that the condition for the existence of a minimum within 0 and 1 is  $S_2 < S_1$ ; this being the case, the further condition that the minimum be negative is equivalent to demanding that  $S_1$  be greater than one—implying a negative fitness coefficient. We conclude that the denominator is always positive for  $p$  lying in the range (0, 1).

Therefore the sign of  $\Delta p$  is determined by

$$-S_2p + f(1 - S_1) - S_1(1 - 2p), \quad (\text{A10})$$

at  $p = 0$ , this is

$$f(1 - S_1) - S_1, \quad (\text{A11})$$

and at  $p = 1$ ,

$$-S_2 + f(1 - S_1) + S_1. \quad (\text{A12})$$

We need (A11) to be  $> 0$  and (A12)  $< 0$  for a polymorphism to exist. This gives condition (3) of the text with the subsidiary requirement  $S_2 > 2S_1$ . Because  $S_2 > 2 \cdot S_1$ ,

$$\frac{1 - S_2}{(1 - S_1)^2} < \frac{1 - 2 \cdot S_1}{(1 - S_1)^2} \leq 1 \text{ trivially.} \quad (\text{A13})$$

In the frequency-dependent case, the conditions—for a polymorphism, for spread to fixation, or for elimination—follow from using  $S_1(p)$  and  $S_2(p)$  in (A11) and (A12) and imposing the appropriate requirements on the sign of  $\Delta p$ .

**References**

- Brookfield J F Y 1982 *J. Theor. Biol.* 94: 281-299  
Chandra H S 1985 *Proc. Natl. Acad. Sci. USA* 82: 1165-1169  
Charlesworth B and Charlesworth D 1983 *Genet. Res.* 42: 1-27  
Crow J F 1979 *Sci. Am.* 240: 134-146  
Crow J F and Kimura M 1970 *An introduction to population genetics theory* (New York: Harper and Row)  
Ginzburg L R, Bingham P M and Yoo S 1984 *Genetics*. 107: 331-341  
Harshey R M and Bukhari A I 1981 *Proc. Natl. Acad. Sci. USA* 78: 1090-1094  
Hiraizumi Y, Sandler L and Crow J F 1960 *Evolution* 14: 433-444  
Hickey D A 1982 *Genetics* 101: 519-531  
Langley C H and Montgomery E A 1983 *Genetics* 104: 473-483  
Langley C H, Brookfield J F Y and Kaplan N L 1983 *Genetics* 104: 457-471  
Simons J and Crow J F 1977 *Annu. Rev. Genet.* 11: 49-78