Transposable element number in mixed mating populations

MARTIN T. MORGAN*
School of Biological Sciences, Washington State University, Pullman, WA 99164–4236, USA
(Received 22 May 2000 and in revised form 9 January 2001)

Summary
Theoretical population genetic studies of transposable elements focus almost exclusively on random mating species, whereas many plants reproduce through partial or substantial self-fertilization. Here I develop computer simulation and analytic approximations of simplified element dynamics (transposition balanced by selective elimination) in partially self-fertilizing populations, using Ty1-copia elements for biological inspiration. Under the most plausible models and parameter values, element numbers decrease with self-fertilization when element insertions are deleterious, but may increase when ectopic exchange regulates element number. Conclusions for models of ectopic exchange depend in part on parameters for which little firm empirical evidence is available. Small changes in selfing rate can lead to abrupt changes in element number when homozygous and heterozygous elements have markedly different fitness effects. Equilibrium element numbers can be sensitive to population size, especially at high selfing rates. Elements are frequently lost in small highly selfing populations under the deleterious insertion model. In contrast, small highly selfing populations can accumulate very large numbers of elements under ectopic exchange. Empirical data on element number and localization in plants with different mating systems suggests that deleterious insertion, rather than ectopic exchange, may regulate element number. Limitations to available empirical data, especially the lack of comparison between closely related species differing in mating system, mean that this conclusion is tentative.

1. Introduction
Mobile genetic elements (e.g. retroposons) are ubiquitous components of plant genomes, as evidenced by PCR methods that detect retroposons in many, diverse plant species (Flavell et al., 1992a; Voytas et al., 1992; Hirochika & Hirochika, 1993; Suonimi et al., 1998; Kumar & Bennetzen, 1999). Retroposons account for considerable amounts of genomic DNA (e.g. > 1/2 the nuclear DNA of maize; SanMiguel et al., 1996), with copy numbers ranging from tens to millions per plant (Flavell et al., 1992a). Evidence indicates that retroposons participate in phenotypic mutation (Berg & Howe, 1989; Bhattacharyya et al., 1990), quantitative and fitness variation (Charlesworth, 1987; Mackay et al., 1992; Keightley, 1994; Lyman et al., 1996), gene regulation (Weil & Wessler, 1990; White et al., 1994; Wessler et al., 1995), deleterious and beneficial mutation (McDonald, 1995), adaptive host evolution (McClintock, 1984; Wilke et al., 1992; White et al., 1994), response to stress (Wessler, 1996; Mhiri et al., 1997; Capy et al., 2000) and genomic structure (Nevers & Saedler, 1977; Bennetzen, 1996; Voytas, 1996; Matzke & Matzke, 1998). Element copy number varies greatly between plant species (Grandbastien, 1992): Arabidopsis has until recently been thought to contain only a few apparently non-functional elements (Voytas et al., 1990), whereas maize and barley may have hundreds of thousands of copies (Pearce et al., 1996b). Such variation in copy number contrasts with the situation in insects and mammals where, with the exception of mammalian LINEs, elements are typically represented by tens of copies (Kumar, 1996). Despite the wealth of empirical data and diversity of selective explanations (Kidwell & Lisch, 2000), few studies develop theoretical expectations for element numbers in species, such as...
Heterozygous elements (outcrossing)  
Zygote  
Ectopic exchange  
Gametes  
Gametes  
Gametes  
Gametes  
Homozygous elements (selfing)  
Zygote  
Retroposon

Fig. 1. Ectopic exchange reduces fitness through production of insertions, deletions and translocations. Heterozygosity may increase ectopic exchange (left) compared with homozygosity (right).

many plants, with reproduction through means other than random mating.

The current population genetic paradigm for understanding element dynamics involves a balance between transposition and selection (Charlesworth & Langley, 1989; Charlesworth et al., 1994). Each element replicates with probability \( u \), and is excised with probability \( v \). In the \( i \)th individual, with \( n \) elements, the rate of transposition is \( (u - v)y_n \). Transposition is therefore density dependent, such that individuals with more elements experience greater numbers of transposition events. Particular insertions may persist in populations because of their adaptive significance, or may be co-opted for adaptation (e.g. McDonald, 1995; Lonning & Saedler, 1997; McFadden & Knowles, 1997). Considerable evidence suggests, however, that element insertions are usually harmful (Charlesworth et al., 1994). There are two popular models of selection against element insertions (Nuzhdin, 1999). Under the ‘deleterious’ model, element insertions disrupt gene function and therefore decrease organism fitness. The ‘ectopic exchange’ model (Montgomery et al., 1987; Langley et al., 1988), on the other hand, envisions unequal crossing over between elements at non-homologous insertion sites, resulting in harmful chromosomal rearrangements (Fig. 1; see Clegg et al., 1997; Caceres et al., 1999). For stable equilibrium element numbers under either selection model, log fitness must decline more than linearly with element number (Charlesworth, 1985; Brookfield, 1986).

Selection acting on transposable elements is expected to be weak, depending on infrequent events, and therefore difficult to observe. Indirect methods for evaluating selection models have therefore been proposed. The deleterious and ectopic exchange models make differing predictions about chromosomal localization of elements. The deleterious mutation model suggests that elements are depauperate in regions of enhanced selection, especially the X chromosome of Drosophila where hemizygous expression exposes heterozygous elements to selection. Ectopic exchange anticipates accumulation of elements in areas of reduced recombination, such as centromeric and telomeric regions. Empirical research supporting the plausibility and predictions of both models is available, but the interpretation of evidence is equivocal (Biemont et al., 1997; Charlesworth et al., 1997; Hooglund et al., 1997).

Many plants reproduce through a mixture of self-fertilization and random mating. Intuitive arguments by Charlesworth & Charlesworth (1995) suggest that differences in mating system may distinguish between the deleterious and ectopic exchange models of element regulation. Self-fertilization and other forms of inbreeding increases homozygosity. Greater homozygosity increases the strength of selection against deleterious insertions. If element number is regulated through deleterious insertion, selfing species may have fewer elements than outcrossing species. On the other hand, homozygosity may decrease the likelihood of ectopic exchange, since elements will usually have homologues to pair with. If element number is regulated through ectopic exchange, selfing species may have more elements than outcrossing species. Wright & Schoen (1999) have developed preliminary simulations of infinite and finite populations undergoing mixed mating that support the intuitive arguments and conclusions of Charlesworth & Charlesworth. Empirical data from Lycopersicon species differing in mating system (Young et al., 1994) may support the role of ectopic exchange in element regulation (Charlesworth & Charlesworth, 1995; although see Francis et al., 1995). With one exception (Wright & Schoen, 1999), formal population genetic treatments of transposable element dynamics focus exclusively on random mating populations.

The objective of this study is to explore the consequences of mating system for equilibrium transposable element number, using the Ty1-copia group of elements as the paradigm. I develop computer simulations and analytical approximations of element transposition and selection under mixed mating. These calculations are used to determine equilibrium element number as a function of mating system, transposition rate and selection model. Simulations are also used to investigate the effects of finite population size. The results show how differing models of element regulation influence equilibrium element number under mating system variation. This provides a context for interpreting the emerging wealth of data on element number and location in outbreeding and inbreeding species.

2. Methods
(i) Transposition, excision and selection
Methods developed here follow element transposition, excision and selection through ‘synergistic’ deleterious
insertion or ‘mass action’ ectopic exchange. Transposition and excision, occurring with probability \( u \) and \( v \) per element, respectively, are assumed to occur independently for each element. If the \( i \)th individual has \( y_i \) heterozygous and \( z_i \) homozygous elements before transposition, the expected number of elements after transposition will be

\[
y'_i = y_i + u(y_i + 2z_i) - v(y_i - 2z_i), \\
z'_i = (1 - 2v)z_i.
\]

(1)

Transposition increases heterozygous element number in proportion to the total number of elements, \( u(y_i + 2z_i) \). Excision reduces heterozygous element number by \( v y_i \) and homozygous number by \( \approx 2v z_i \), but each excision at a homozygous site creates a heterozygous insertion. Thus change in heterozygous elements due to excision is \(-v y_i + 2v z_i = -v(y_i - 2z_i)\). Estimates from *Drosophila* suggest that \( u \) varies between 0.00001 and 0.001 (Nuzhdin & Mackay, 1994, 1995), with \( v \approx u/10 \). Estimates of transposition rates in natural plant populations are not available. Low transposition rates (e.g. \( u = 0.0001 \)) result in large stochastic variation in even moderate-sized populations (Charlesworth & Charlesworth, 1983; Tsitrono et al., 1999). For tractability, values of \( u = 0.001 \), \( v = 0.0001 \) are used in most simulations. No attempt is made to include a faster than linear increase in transposition rate with element number, for which there is evidence in *Drosophila* lines with unusually high rates of movement of *copia* (Nuzhdin et al., 1997).

The ‘synergistic’ selection model assumes that element insertions are deleterious, with effects of insertions interacting synergistically to influence individual fitness. Under this model, the fitness of an individual with \( y_i \) heterozygous and \( z_i \) homozygous elements is

\[
w(y_i, z_i) = \exp\{-\alpha(h y_i + z_i) - \beta(h y_i + z_i)^2/2\}. \tag{2}
\]

Heterozygous and homozygous elements influence fitness differently, as determined by the dominance coefficient \( h \). The selection model is analogous to that used for multilocus deleterious mutation (Kondrashov, 1984, 1985; Charlesworth et al., 1991; Crow, 1993), and the dominance coefficient is usually \( h \approx 0.5 \). Estimates of \( \alpha \) and \( \beta \) are unavailable, but parameter values can be chosen to agree with empirical observation of element number in *Drosophila* and plausible parameters of selection on deleterious mutation (although see Charlesworth, 1998). Given transposition rates of \( u = 0.001 \) and dominance coefficient \( h = 0.2 \), equilibrium element copy number in random mating populations of \( \tilde{y} \approx 50 \) results when \( \alpha \approx 8.33 \times 10^{-4}, \beta = \alpha/2 \).

The ‘mass action’ model of ectopic exchange assumes that increasing element number increases opportunity for exchange between non-homologous insertion sites. One version of fitness in this model is

\[
w(y_i, z_i) = \exp\{-\omega^2(ay_i + bz_i)^2/2\}. \tag{3}
\]

Fitness decreases with the square of element number, and with the rate of recombination in the genome. The dependence of fitness on element number is based on the principle that the harmful effects of ectopic exchange require participation of two elements, chosen at random and hence influencing fitness through the square of element number. Evidence for ectopic exchange is available in *Drosophila* (e.g. Montgomery et al., 1991; Lim & Simmons, 1994; Caceres et al., 1999) and Ac element-induced ectopic exchange is used as an experimental tool in, for instance, maize (Shalev & Levy, 1997). The mass action model of ectopic exchange assumes that chromosomal proximity of elements does not influence ectopic exchange, whereas ectopic exchange is thought to occur more commonly between elements in physical proximity (Lim & Simmons, 1994). Such localized exchange would tend to distribute elements evenly over the chromosome, and the essential feature of the mass action model (stronger selection with increasing element number) remains relevant. Equation (3) is motivated by biological realism, but relies on simplification for modelling convenience. Parameters \( a \) and \( b \) reflect the opportunity for ectopic exchange between heterozygous and homozygous insertions. The *a priori* expectation is that \( b < a \), since elements at homozygous insertion sites may be less prone to non-homologous pairing. There is no direct empirical evidence for decreased ectopic exchange with increased homozygosity in plants, so various parameter values of \( b \) are investigated. The parameter \( \omega^2 \) reflects the frequency with which non-homologous recombination occurs. As with \( \alpha \) and \( \beta \) in (2), \( \omega^2 \) can be chosen to agree with empirically observed element numbers. Models below use \( \omega = 1 \times 10^{-3} \), resulting in equilibrium element number in outcrossing species of \( \approx 50 \) when \( a = 1 \).

(ii) Simulation

Three different approaches were taken to modelling the evolutionary dynamics of transposable element evolution. The first involves detailed simulation, following chromosome insertion sites in a genome. The second approach again uses simulation, but assumes that the details of element insertion location do not matter for element evolution. Finally, analytical approximations follow mean element number in cohorts of individuals.

Computer simulations were developed to follow individual elements inserted at specific chromosome locations. The simulations model distinct individuals, each individual with several chromosomes with nu-
merous sites for element insertion. Chromosomes of each individual are represented as strings of binary numbers, with '0' indicating no element inserted and '1' indicating element insertion at specific locations. Individuals consist of several chromosomes, in addition to summary information about the number of heterozygous and homozygous element \((y_i, z_i)\). In the simulations reported below, individuals consist of 7 chromosomes, each experiencing an average of 1 recombination event per chromosome.

Elements are followed through stochastic transposition, mating and selection. Transposition is determined for each individual. The number of transpositions and excisions is determined by drawing a Poisson deviate from a distribution with mean defined in (1). Each new transposition is then located in the genome by drawing a uniform deviate. Insertions into sites already occupied were ignored, since such insertions are rare except when element number becomes large.

Mating starts with choice, using a uniform random deviate, of an individual to act as female parent. A second uniform deviate determines mode of reproduction (self-fertilization with probability \(S\), random mating with probability \(1 - S\)). Suppose the offspring is to be produced through self-fertilization. Recombination sites on each chromosome are generated in two steps. The number of recombination events is determined by drawing a Poisson deviate with mean \(r\). The \(r\) locations of the recombination events are then placed on the chromosome using uniform deviates. Gametes are then constructed using the pattern of inheritance implied by the recombination map. Computationally, these operations are carried out using bitwise logical operations (Fraser & Burnell, 1970; Charlesworth et al., 1992). A second gamete is constructed using similar steps. Offspring produced through outcrossing are constructed in the same way, except that the second gamete comes from another individual chosen by drawing a uniform random deviate. The fitness of the offspring, calculated according to (2) or (3), is compared with a uniform deviate. If the fitness is less than the deviate, the offspring is rejected and the process of choosing parents and constructing gametes repeated in its entirety. This continues until \(N\) offspring are obtained.

Simulations reported below follow \(N = 10000\) individuals. Populations are started with Poisson-distributed heterozygous elements, with mean equal to 50. Simulations run for 5000 generations to establish an approximate equilibrium. The population is then censused every 100 generations, with the final census at generation 15000. Most parameter combinations require several hours of computer time.

(iii) ‘Fast’ simulation

‘Fast’ simulations were developed to follow only the number, rather than the location, of elements in each individual. The \(i\)th individual in the simulation is characterized by the number of heterozygous and homozygous elements \((y_i, z_i)\). Individuals are followed through stochastic transposition, mating and selection. Transposition and excision are simulated by drawing a Poisson deviate, as in (1).

Choice of parents for mating follows the procedure outlined for the full simulation. Suppose the offspring is to be produced through self-fertilization. The number of heterozygous elements in the offspring, \(y''_i\), is determined by drawing a binomial deviate from a sample of size \(y'_i\) with probability \(1/2\). The number of homozygous elements, \(z''_i\), is then chosen by drawing a binomial deviate from a sample of size \(y'_i - y''_i\) with probability \(1/2\), and adding this to the number of homozygous elements in the parent, \(z'_i\). Suppose instead that the offspring is produced through random mating. The number of elements in the female gamete equals \(z'_i\) plus a random deviate drawn from a sample of size \(y'_i\) with probability \(1/2\). The number of heterozygous elements in the outcrossed offspring, \(y''_i\), is determined by adding numbers of elements in male and female gametes. Homozygous element number is \(z''_i = 0\). This assumes that insertions at a particular site are rare so that random mating does not result in homozygous elements. Fitnesses of offspring are calculated, and offspring accepted or rejected, as outlined in the full simulation.

Simulations reported below follow \(N = 10000\) individuals. Populations are started with Poisson-distributed heterozygous elements, with mean equal to 50. Simulations run for 5000 generations to establish an approximate equilibrium. The population is then censused every 100 generations, with the final census at generation 15000. Most parameter combinations require several hours of computer time.

(iv) Analytical approximation

The analytical approximation follows mean heterozygous and homozygous element number in cohorts \((y_i, z_i)\). Cohorts \(i = 0, 1, 2, \ldots, n\) correspond to subsets of the population resulting from \(i\) generations of self-fertilization. The frequency of the \(i\)th cohort is written as \(p_i\).

Transposition in each cohort changes heterozygous element number deterministically: \(y'_i = y_i + u(y_i + 2z_i) - v(y_i - z_i)\) and \(z'_i = (1 - 2v)z_i\). Transposition does not change cohort frequency, so \(p'_i = p_i\).

Reproduction occurs through self-fertilization, probability \(S\), or random mating, probability \(1 - S\). The transition in element number associated with self-fertilization is \(y_{i+1}' = y_i'/2\) and \(z_{i+1}' = z_i' + y_i'/4\). The
frequency of the $(i+1)$th cohort after self-fertilization is $p_{i+1}^x = Sp_i$. The mean element number in the $n$th (i.e. final) cohort is calculated as the average of the $n$th and $(n-1)$th cohorts, $p_n^x = S(p_{n+1}^x + p_{n-1}^x)$. Random mating determines element number in cohort 0. Random mating is assumed to result only in heterozygous elements (i.e. selection is sufficiently strong that elements are individually rare). The number of heterozygous elements is calculated as the average number of heterozygous elements in gametes contributed by each cohort. Thus $\bar{y} = \sum_{i=0}^{n} p_i (\bar{y}_i / 2 + \bar{z}_i)$ and $\bar{z}_0 = 0$. The frequency of the 0th cohort after random mating is $p_0^x = (1 - S)$.

Change in element number due to selection is based on a quantitative genetic approximation. The change in the mean of trait $x$ due to selection is $\Delta x = \sigma^2 x \ln \bar{w} / \ln \bar{x}$ (Lande, 1976; Falconer & Mackay, 1996). The variance of $x$ is $\sigma^2_x$, and it is assumed that trait $x$ is uncorrelated with other traits. Selection is applied within each cohort, with cohort mean fitness, $\bar{w}_i$, changing the frequency of the cohort, $p_i = p_i^x \bar{w}_i / \bar{w}$. Cohort mean fitness is approximated as the fitness at the cohort average element number, $\bar{w}_i \approx w(\bar{y}_i, \bar{z}_i)$. Several authors discuss conditions when this approximation is accurate (e.g. Abrams et al., 1993). Specific fitness functions are given above for synergistic (equation 2) and mass action (equation 3) models of selection. Within each cohort, the distribution of heterozygous and homozygous elements is assumed to follow a bivariate Poisson distribution. The variance of heterozygous and homozygous element numbers is then $(\bar{y}_i^2, \bar{z}_i^2)$. It is assumed that there is no correlation between heterozygous and homozygous elements between individuals within each cohort. Justification for these assumptions is threefold. Theoretical arguments suggest that element number should be approximately Poisson-distributed in random mating populations (Langley et al., 1983). Quantitative genetic studies of selection in mixed mating species have successfully applied a cohort approach (Kelly, 1999a, b), partly because identity disequilibrium (cf. Lande et al., 1994) generated by self-fertilization is likely to exceed greatly that generated by selection. Finally, tedious calculation following both mean and variance of element numbers within cohorts (using equations modified from Charlesworth et al., 1991) lead to results very similar to those obtained here (M. T. Morgan, unpublished.).

The foregoing outlines how transposition, mating and selection influence element number over a single generation. Roots of the resulting non-linear, complicated set of equations can be solved explicitly. Here, however, equations were iterated until proportional change in average element number between generations was less than $10^{-6}$. The number of cohorts followed was set to 20, although variation in cohort number greater than $\approx 10$ has little effect on results obtained. The convergence criterion is usually satisfied after a few seconds of computer time.

3. Results

(i) Synergistic deleterious insertion

Fig. 2 shows equilibrium element number under synergistic selection. When net transposition, $u = v$, equals 0.001, greater self-fertilization reduces equilibrium element number. Reduction in element number makes intuitive sense, because self-fertilization increases the number of homozygous relative to heterozygous elements, and hence the strength of selection against elements (Kimura & Maruyama, 1966). In addition, self-fertilization increases the variance in element number between individuals, allowing greater opportunity for selection (Burt & Trivers, 1998). Complete self-fertilization reduces element number more than 50-fold. This dramatic reduction reflects the feedback effect of reduced element number on rate of transposition: stronger selection reduces element number, effectively decreasing the number of elements undergoing transposition each iteration. Analytical approximation agrees closely with both the full and fast simulations when selfing rate is less than 0.8 (Fig. 2, open symbols and lines). Discrepancies at high selfing rate occur, when all elements are lost from the simulations. Apparently, equilibrium element number decreases sufficiently that genetic drift can result in fixation of element-free genotypes, even in large populations. The effects of population size are discussed in greater detail below.

Qualitatively different patterns of element number occur when transposition rate becomes large (e.g. $u = 0.01$ in Fig. 2). Element number remains high and approximately constant until intermediate selfing ($S \approx 0.62$), and then decreases rapidly. These analytical results are supported by fast simulations. Unfortunately, the large number of elements involved precludes full simulation of this effect. A similar threshold is observed in models of deleterious mutation, provided that the genomic mutation rate is large relative to the dominance coefficient (Lande et al., 1994). The heuristic explanation for this pattern is that weak selection allows heterozygous elements to accumulate to high levels. Selfed progeny then have many homozygous loci and extremely low fitness. The post-selection population consists almost exclusively of outcrossed individuals. Under these circumstances, homozygosity associated with self-fertilization is ineffective at reducing average element number.

Fig. 3 investigates the threshold effect in more detail, using results from the analytical model. An abrupt threshold like that in Fig. 2 is observed for all net rates of transposition. There is a direct relationship
between transposition rate and dominance coefficient, required for the threshold to appear. For instance, when \( u-v = 0.01 \) a dominance coefficient of \( h = 0.2 \) generates a threshold (Fig. 2), while \( u-v = 0.0005 \) exhibits a threshold only at \( h \leq 0.05 \) (Fig. 3). Increasing the transposition rate or decreasing the dominance coefficient increases the selfing rate at which the threshold occurs (Fig. 3).

(ii) **Mass action ectopic exchange**

Fig. 4 shows equilibrium element number under the mass action model of ectopic exchange. When element insertions are additive at individual sites \( (a = 2b) \), self-fertilizing populations have half as many elements as outcrossed populations. This makes intuitive sense, as each selective elimination removes two elements in the selfer rather than one in the outcrosser (Kimura & Maruyama, 1966). It is more biologically plausible to assume that homozygous elements are less prone to ectopic exchange than heterozygous elements \( (a \geq b) \). Self-fertilization then results in an increase in element number. The effect tends to be gradual over most of the range of selfing rates. Decreasing effects of homozygous insertions delays increase in element number until a higher selfing rate, but the increase becomes more dramatic. Simulation and analytical calculations are in agreement when selfing rate is less
than $\approx 0.8$. At high selfing rates and when $a \gg b$, element number increases to the computational limits of simulation. Stochastic effects at high selfing rate are discussed further below.

With small values of $b$, there is a threshold complementary to that observed under synergistic selection with low dominance coefficients. The reason for this is comparable to the selective interference under synergistic selection, but with self-fertilized (homozygous) individuals having high fitness compared with outcrossed (heterozygous) individuals. Additional results show that the occurrence and location of the threshold is influenced by transposition rate and relative selective effects on heterozygous and homozygous elements: the threshold occurs at higher selfing rate when $b$ decreases or net transposition increases.

(iii) **Mean fitness and modifier selection**

Fig. 5 depicts the relationship between population mean fitness and rate of net transposition under different mating systems. For all models of selection, greater net transposition decreases mean fitness. Mating system does not affect the decrease in mean fitness under mass action ectopic exchange when $b = 1$ (lower panel; $b = 1$ corresponds to equal occurrence of ectopic exchange at heterozygous and homozygous sites). With synergistic selection (upper panel) and with ectopic exchange occurring more frequently in heterozygotes than homozygotes (lower panel, $b = 0.5$), mean fitness decreases much more rapidly in outcrossing than self-fertilizing populations.

The slope of the relationship between the logarithm of mean fitness and rate of net transposition, $d \ln w / d(u-v)$, is the selection gradient acting on individual net transposition. Under synergistic selection models, selection for reduced net transposition is stronger in outcrossing populations than in self-fertilizing populations. With mass action ectopic exchange, fitness parameters resulting in greater element number (i.e. $b < 1$, Fig. 4) enhance selection against net transposition much more strongly in self-fertilizing than in outcrossing populations. Evolutionary modification of the net transposition rate therefore obscures the mating system differences in element number outlined above.

(iv) **Population size**

Fast simulations show that the usual consequence of finite size under synergistic selection is element loss (Table 1). Simulations varying the rate of excision $u$ (results not shown) indicate that the loss of elements occurs primarily through stochastic fixation of element-free sites, rather than element excision. Element loss occurs more than half the time in very small ($N = 50$) outcrossing populations. The threshold for element persistence when $S > 0.5$ can be large; populations selfing at $S = 0.9$ must be larger than $N = 10000$ individuals for elements to persist in at least half the replicate populations. This probably reflects the low equilibrium frequency expected under self-fertilization, with small fluctuations in average element number sufficient to result in element loss.

Table 2 documents unusual behaviour in small populations under synergistic selection with complete self-fertilization ($S = 1.0$). While elements are lost from 68–92% of replicates (Table 1), elements increase to substantial frequency in the remaining simulations. Long-term average element number in those simulations where elements persist decrease with increasing
Entries have the form \( A/B \), where \( A \) is the percentage of simulations in which elements were lost and \( B \) is the percentage of simulations where average element number exceeds 1000. Italicized results are discussed in the text and Table 2.

### Table 2. Unusual equilibrium numbers of heterozygous, \( y \), and homozygous, \( z \), elements in small populations at high selfing rates (\( S = 1.0 \)) under synergistic selection

<table>
<thead>
<tr>
<th>Population size, ( N )</th>
<th>( \bar{y} )</th>
<th>( \bar{z} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.73</td>
<td>150</td>
</tr>
<tr>
<td>100</td>
<td>0.34</td>
<td>70.9</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>32.3</td>
</tr>
<tr>
<td>500</td>
<td>0.035</td>
<td>7.25</td>
</tr>
</tbody>
</table>

Table 3. Influence of population size and selfing rate on element fate under mass action ectopic exchange.

<table>
<thead>
<tr>
<th>Population size, ( N )</th>
<th>Rate of self-fertilization, ( S )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0·0</td>
</tr>
<tr>
<td>50</td>
<td>46/0</td>
</tr>
<tr>
<td>100</td>
<td>12/0</td>
</tr>
<tr>
<td>200</td>
<td>0/0</td>
</tr>
<tr>
<td>500</td>
<td>0/0</td>
</tr>
<tr>
<td>1000</td>
<td>0/0</td>
</tr>
<tr>
<td>5000</td>
<td>0/0</td>
</tr>
<tr>
<td>10000</td>
<td>0/0</td>
</tr>
</tbody>
</table>

Entries have the form \( A/B \), where \( A \) is the percentage of simulations in which elements were lost and \( B \) is the percentage of simulations where average element number exceeds 1000.
element number at $S \approx 0.8$ in Fig. 4. The high number of homozygous elements means that outcrossed progeny have very low fitness compared with selfed progeny, so the population consists of highly inbred lineages. Element number continues to increase rapidly once inbred lineages become established. The rapid increase reflects the positive feedback of transposition, with greater element number facilitating increased numbers of new transpositions. Element number appears to increase without limit ($\bar{n} > 4000$) in smaller highly selfing populations.

4. Discussion

The models of element regulation investigated here predict different patterns of element number in self-fertilizing compared with outcrossing species. When element number is regulated by the deleterious effects of insertion (synergistic selection), increasing self-fertilization decreases element number. Weak dominance results in a threshold effect where element number only declines above an intermediate selfing rate. In many predominantly selfing populations, genetic drift is likely to result in loss of all elements. Element number may, however, increase in highly selfing populations ($S > 0.95$). When element number is regulated through ectopic exchange, and when homozygous elements are less likely to exhibit ectopic exchange than heterozygous elements ($b \leq 1$, Fig. 1), self-fertilization increases equilibrium element number. The increase is usually gradual, but can be abrupt in small populations or at high selfing rates and when homozygous insertions are much less likely than heterozygous insertions to participate in ectopic exchange. Accumulation of element number under ectopic exchange in self-fertilizing species is accompanied by increasingly strong selection for decreased net transposition (Fig. 5).

The following summarizes existing empirical studies on element number and localization. The summary provides some support for the importance of deleterious insertion, rather than ectopic exchange, in regulating element number.

(i) Element number

Table 4 summarizes empirical estimates of Ty1-copia element number in relation to mating system. The Ty1-copia element family is the best characterized of plant transposable elements. Most studies estimating element number investigate economically important or model species. Element number is usually assessed using heterologous probes or primers. This results in qualitative estimates, especially when element number exceeds $\approx 100$. Mating system characterization is from Tables 1.4 and 1.5 of Frankel & Galun (1977). Outbreeding taxa typically have many elements ($\geq 10000$). Inbreeding taxa frequently have fewer elements ($\leq 1000$), with prominent exceptions (Hordeum vulgare, Oryza australiensis, Vicia faba). This pattern is more consistent with deleterious insertion, rather than ectopic exchange, but statistical support for the pattern is weak. Non-parametric analysis, excluding species with only qualitative element number estimates, results in marginally significant differences (Kolmogorov–Smirnov test, $p \approx 0.044$); differences are not significant using a parametric $t$-test. The statistical analysis is extremely

![Fig. 5. Relationship between mean fitness (logarithmic scale) and net rate of transposition ($u−v; u = 10v$) under synergistic selection ($\alpha = 8.3\times10^{-4}, \beta = \alpha/2, h = 0.2$) and mass action ectopic exchange ($a = 1, \omega = 1 \times 10^{-3}$, values of $b$ as indicated). ‘Selfing’ values correspond to $S = 0.99$, ‘outcrossing’ values to $S = 0.01$. Mean fitness does not depend on mating system when $b = 1$.](https://www.cambridge.org/core/terms)
Table 4. Empirical relationship between mating system and Ty1-copia-like element number

<table>
<thead>
<tr>
<th>Species</th>
<th>Number</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inbreeding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>200</td>
<td>Brandes et al., 1997; &lt; 3 per family, 28 families reported by Voytas &amp; Ausubel, 1988; Voytas et al., 1990; Konieczny et al., 1991; Wright et al., 1996; ‘several’ copies from 100 families, Kapitonov &amp; Jurka, 1999; 14% of genomic sequence, Ty1-copia fraction not specified, The Arabidopsis Genome Initiative, 2000</td>
</tr>
<tr>
<td>Cicer arietinum</td>
<td>600</td>
<td>Sant et al., 2000</td>
</tr>
<tr>
<td>?C. reticulatum</td>
<td>10</td>
<td>Sant et al., 2000</td>
</tr>
<tr>
<td>Citrus spp.</td>
<td>Multiple</td>
<td>Asins et al., 1999</td>
</tr>
<tr>
<td>Glycine max</td>
<td>500–800</td>
<td>Luten &amp; Morris, 1993</td>
</tr>
<tr>
<td>Gossypium spp.</td>
<td>100s–1000s</td>
<td>VanderWiel et al., 1993</td>
</tr>
<tr>
<td>Hordeum vulgare</td>
<td>≈ 70000</td>
<td>Suoniemi et al., 1996; Vicient et al., 1999</td>
</tr>
<tr>
<td>?Lycopersicon chilens</td>
<td>2200</td>
<td>Yanez et al., 1998</td>
</tr>
<tr>
<td>L. esculentum</td>
<td>200</td>
<td>Unpublished, in Kumar et al., 1997</td>
</tr>
<tr>
<td>Nicotiana tabacum</td>
<td>&gt; 100</td>
<td>Grandbastien et al., 1989; Pouteau et al., 1991</td>
</tr>
<tr>
<td>&gt; 30</td>
<td></td>
<td>Tto1, Hirochika, 1993</td>
</tr>
<tr>
<td>Oryza spp.</td>
<td>1000</td>
<td>Hirochika et al., 1992</td>
</tr>
<tr>
<td>&gt; 1–5</td>
<td></td>
<td>ToS17, Hirochika et al., 1992</td>
</tr>
<tr>
<td>O. sativa</td>
<td>Moderate/high</td>
<td>Wang et al., 1997; &lt; 1-5% of genome, Mao et al., 2000</td>
</tr>
<tr>
<td>? O. australiensis</td>
<td>Extraordinary</td>
<td>Nakajima et al., 1996; Noma et al., 1997</td>
</tr>
<tr>
<td>Phaseolus vulgaris</td>
<td>40</td>
<td>Garber et al., 1999</td>
</tr>
<tr>
<td>Pisum sativum</td>
<td>50</td>
<td>Grandbastien, 1992</td>
</tr>
<tr>
<td>Solanum tuberosum</td>
<td>&gt; 360</td>
<td>Camirand et al., 1990; Flavell et al., 1992b</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>200</td>
<td>Harberd et al., 1987; Moore et al., 1991</td>
</tr>
<tr>
<td>Vicia faba</td>
<td>1000000</td>
<td>Pearce et al., 1996a</td>
</tr>
<tr>
<td>?V. melanolps</td>
<td>1000</td>
<td>Pearce et al., 1996a</td>
</tr>
<tr>
<td>V. sativa</td>
<td>5000</td>
<td>Pearce et al., 1996a</td>
</tr>
<tr>
<td>Vitis spp.</td>
<td>&gt; 20</td>
<td>Verries et al., 2000</td>
</tr>
<tr>
<td>Outbreeding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allium cepa</td>
<td>10000</td>
<td>Pearce et al., 1996b</td>
</tr>
<tr>
<td>Beta vulgaris</td>
<td>High</td>
<td>Schmidt et al., 1995</td>
</tr>
<tr>
<td>Elaeis guineensis</td>
<td>‘Relatively low’</td>
<td>Castilho et al., 2000</td>
</tr>
<tr>
<td>Prunus spp.</td>
<td>1–3</td>
<td>Asins et al., 1999</td>
</tr>
<tr>
<td>Pinus spp.</td>
<td>High</td>
<td>Kamm et al., 1996</td>
</tr>
<tr>
<td>Secale cereale</td>
<td>1000000</td>
<td>Pearce et al., 1997</td>
</tr>
<tr>
<td>Zea maize</td>
<td>&gt; 30000</td>
<td>Opie 1, SanMiguel et al., 1996</td>
</tr>
<tr>
<td>&gt; 10000</td>
<td></td>
<td>PREM-2, Turcich et al., 1996</td>
</tr>
<tr>
<td>1–5</td>
<td></td>
<td>Bsl1, Jin &amp; Bennetzen, 1989</td>
</tr>
</tbody>
</table>

Mating system characterized by Frankel & Galun (1977). The mating system of taxa preceded by a question mark (?) is uncertain.

Sensitive to the necessarily approximate counts from some species, and to inclusion of extreme counts in either mating system category.

There are several obvious limitations to the broad-scale comparative study in Table 4. Characterization of mating system is usually qualitative and historically recent, rather than a quantitative estimate of the evolutionarily relevant (i.e. on the timescale during which element number evolves) rate of self-fertilization. For instance, the large showy flowers and natural history of Gossypium suggest substantial outcrossing (Fryxell, 1979), whereas cultivated cotton is characterized as inbreeding (Frankel & Galun, 1977). Exceptional element numbers in H. vulgare, O. australiensis and V. faba might reflect high selling or reductions in population size during domestication (see Table 1). The largely qualitative assessment of element number may also obscure an underlying pattern. Most studies identify short homologous portions of single genes. This does not allow distinction between intact, active elements and fragmented or otherwise inactive elements. In chromosome 2 of Arabidopsis thaliana, for instance, over 50% of transposon sequences are pseudogenes (Lin et al., 1999). Genome size varies amongst species, potentially altering the number of suitable insertion sites (Kumar et al., 1997). In particular, many of the species in Table 4 are polyploid (VanderWiel et al., 1993). Consequences of polyploidy are not investigated in the simulations reported here. Finally, domestication...
may introduce features, such as population bottlenecks, that influence element numbers in unexpected ways.

The theory developed here provides motivation for surveying element numbers across related taxa differing in mating system. Charlesworth & Charlesworth (1995) suggested that element number in *Lycopersicon* spp. (Young et al., 1994) is consistent with ectopic exchange. This is based on the presence of fewer *Lyt1*-hybridizing sequences in green-fruited (putatively outcrossing) species compared with red-fruited (putatively self-fertilizing) species. Green-fruited species appear to contain 2–8 copies, compared with 40–50 in red-fruited species. This is consistent with much less ectopic exchange between homozygous compared with heterozygous elements (e.g. *b < 0.5* in Fig. 4). Rigorous conclusions are undermined by uncertainty surrounding the evolutionarily relevant selling rates of these species (Francis et al., 1995). Candidates for additional study identified in Table 4 include *Vicia* spp., *Pinus* spp. and *Oryza* spp. The contrast between inbreeding *Arabidopsis* and outbreeding *Arabis* is also relevant.

(ii) **Chromosomal localization**

Several authors use chromosomal localization to distinguish between theories of element regulation (Hoogland & Biemont, 1996; Biemont et al., 1997; Charlesworth et al., 1997; Duret et al., 2000). In random mating populations under ectopic exchange, element abundance decreases with greater rates of recombination (Langley et al., 1988; Charlesworth & Langley, 1989), including on the X chromosome where recombination in females is possible (Langley et al., 1988). This assumes that ectopic and meiotic exchange have similar densities, since patterns of ectopic exchange have not been directly documented. Under deleterious insertion, elements are underrepresented in regions where selection is strong (i.e., regions of high gene density). Both theories anticipate lower element abundance in regions of enhanced recombination due to the Hill–Robertson effect (Duret et al., 2000). These predictions change under self-fertilization. Homozygosity reduces opportunity for ectopic exchange (provided *b < 1*), regardless of the rate of recombination in surrounding chromosomal regions. Elements may therefore be dispersed throughout chromosomes of inbreeding species when element regulation occurs through ectopic exchange. In contrast, homzygosity associated with inbreeding may amplify the deleterious consequences of element insertion into functional genes. Elements may therefore be restricted to genomic regions of inbreeding species with few active genes when element regulation occurs through deleterious insertion. Self-fertilization effectively reduces recombination, greatly reducing the importance of the Hill–Robertson effect in influencing chromosomal localization.

Chromosomal location of *Ty1-copia* elements in plants has been studied primarily using *in situ* hybridization (Table 5). *Ty1-copia* elements are

### Table 5. Chromosomal location of *Ty1-copia* elements, determined using *in situ* hybridization

<table>
<thead>
<tr>
<th>Species</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inbreeding</strong></td>
<td></td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>Pericentromeric heterochromatin (Brandes et al., 1997)</td>
</tr>
<tr>
<td><em>Avena</em> spp.</td>
<td>Dispersed. –, Centromeric and nucleolar organizer regions (Katsiotis et al., 1996)</td>
</tr>
<tr>
<td><em>Cicer arietinum</em></td>
<td>Pericentromeric heterochromatin (Brandes et al., 1997)</td>
</tr>
<tr>
<td><em>Hordeum vulgare</em></td>
<td>Dispersed. –, Centromeric regions, telomeres and nucleolar organizer regions (Suoniemi et al., 1996)</td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>Dispersed (Mao et al., 2000)</td>
</tr>
<tr>
<td><em>Sorghum vulgare</em></td>
<td>Dispersed (Millet et al., 1998)</td>
</tr>
<tr>
<td><em>Vicia faba, V. narbonensis</em></td>
<td>Dispersed. –, Centromeric, nucleolar organizing, and tandem repeat regions.</td>
</tr>
<tr>
<td></td>
<td>+, Local (Brandes et al., 1997)</td>
</tr>
<tr>
<td><strong>Outbreeding</strong></td>
<td></td>
</tr>
<tr>
<td><em>Allium cepa</em></td>
<td>Dispersed. +, Terminal heterochromatin (Pearce et al., 1996b)</td>
</tr>
<tr>
<td><em>Brassica campestris, B. oleracea</em></td>
<td>Dispersed. –, Centromeric, nucleolar organizing and tandem repeat regions.</td>
</tr>
<tr>
<td></td>
<td>+, Local (Brandes et al., 1997)</td>
</tr>
<tr>
<td><em>Beta vulgaris</em></td>
<td>Dispersed. –, Local (Schmidt et al., 1995)</td>
</tr>
<tr>
<td><em>Pennisetum glaucum</em></td>
<td>Dispersed. –, Centromeric, nucleolar organizing and tandem repeat regions.</td>
</tr>
<tr>
<td></td>
<td>+, Local (Brandes et al., 1997)</td>
</tr>
<tr>
<td><em>Pinus elliottii</em></td>
<td>Dispersed. –, Centromeric, 18S, 5S, 25S rRNA (Kamm et al., 1996)</td>
</tr>
<tr>
<td><em>Secale cereale</em></td>
<td>Dispersed. –, Centromeric regions and terminal heterochromatin</td>
</tr>
<tr>
<td></td>
<td>(Pearce et al., 1997)</td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>Dispersed (Millet et al., 1998)</td>
</tr>
</tbody>
</table>

Mating system characterized by Frankel and Galun (1977). +/− indicates regions where element numbers are enriched or reduced.
dispersed throughout chromosomes in all species investigated, except for *A. thaliana* and the inbreeding *Cicer arietinum*. Elements are often absent or at reduced frequency in centromeric, pericentromeric or nucleolar organizing regions, and occasionally enriched in pericentromeric or terminal heterochromatin. There is no obvious patterning of element localization with mating system, with the exception of restricted distribution in inbreeding *A. thaliana* and *C. arietinum*.

Genome sequencing is providing an exquisite look at element localization in the inbreeding *A. thaliana*. Transposable elements are strongly clustered in the pericentromeric region of each chromosome and in regions of low gene expression (The Arabidopsis Genome Initiative, 2000). Clustering of elements near the centromere has been taken as evidence for the importance of ectopic exchange, because recombination is often thought to be reduced in this region (Langley et al., 1988; Charlesworth & Langley, 1989). Although recombination is severely reduced at the centromere, recombination in the pericentromeric regions is suppressed only modestly in *A. thaliana* (Copenhaver et al., 1999; Lin et al., 1999). In addition, as argued above, restricted recombination is less relevant for ectopic exchange (when \( b < 1 \)) in a highly inbreeding species such as *A. thaliana*. The abundance of transposable elements in regions of low gene density may, therefore, represent preferential persistence of insertions in non-coding sequence. Duret et al. (2000) have recently shown that transposons but not retroposons occur in regions of high recombination in self-fertilizing *Caenorhabditis elegans*. These authors argue that this pattern is inconsistent with both ecotopic exchange and deleterious insertion models of element regulation, because the Hill–Robertson effect anticipates element accumulation in areas of restricted recombination. As outlined above, though, self-fertilization eliminates much of the opportunity for chromosomal localization via the Hill–Robertson effect.

### (iii) Summary and future directions

Computer simulations support previous qualitative predictions about how mating system influences element number under different modes of regulation. Broadly speaking, greater self-fertilization decreases element number under deleterious insertion and increases element number with regulation through ectopic exchange. Empirical data, though far from satisfactory, are more consistent with the deleterious effects of insertion than with ectopic exchange. Specifically, element number in inbreeding species is generally reduced compared with outbreeding species. Elements are generally dispersed throughout chromosomes, rather than being restricted to areas of restricted recombination as anticipated by regulation through ectopic exchange. In inbreeding *A. thaliana*, element localization to the pericentromeric region probably reflects the decreased density of coding genes in this region. Unfortunately, this is consistent with both deleterious insertion and ectopic exchange. The simplistic nature of the population genetic models presented here, and the correspondingly crude characterization of element number and localization, emphasize the opportunity for additional theoretical and empirical research into factors regulating element number.

This work received partial support from the US National Science Foundation grant DEB-9974017.

### References


Charlesworth, B. (1985). The population genetics of


