Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 1655–1659, March 1983
Evolution

Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*

(transposable elements/non-Mendelian inheritance/geographic variation/speciation)

MARGARET G. KIDWELL

Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912

Communicated by James F. Crow, December 17, 1982

ABSTRACT **Hybrid dysgenesis is manifested as a group of correlated aberrant genetic traits such as sterility, increased mutation rate, and male recombination. Previous work has shown that it appears when males of strains carrying either of two independent families of transposable elements called *I* and *P* factors are hybridized with females of susceptible strains called *R* and *M*, respectively. Here the results of an extensive survey for dysgenic potential in *Drosophila melanogaster* strains are reported. Striking temporal trends in the distribution of strains were observed with respect to the two transposable element systems; in particular, the frequency of *R* and *M* strains is positively correlated with laboratory age. In recent tests of strain samples, those collected from nature about 50 years ago were the earliest observed to possess *I* characteristics. The *I* type was increasingly frequent in samples from strains more recently originating in the wild. This type is apparently ubiquitous in present day natural populations. The *P* type was not found in strain samples collected before 1950, and collections made subsequently showed increasing frequencies of *P*-factor activity with decreasing laboratory age. Marked geographical patterns are documented in the contemporary worldwide distribution of variant strains within the *P–M* system. *M* strains are currently fairly common in natural populations from various parts of the world, except on the American continent where they are rare. The degree and distribution of quantitative variation within *M* and *P* strain categories is related to their time of origin in the wild. The implications of these results are discussed in relation to the hypothesis that hybrid dysgenesis determinants have evolved recently in natural populations and to an alternative hypothesis of laboratory evolution.**

Approximately 9% of the *Drosophila melanogaster* genome exists as dispersed moderately repetitive sequences (1). These sequences have been grouped into classes based primarily on their internal structure. Two clearly distinct structural classes of middle repetitive sequences are the *copia*-like elements and the foldback elements. A third class is that constituting the *P* elements (2, 3). This is a family of mobile dispersed genetic elements implicated in the *P–M* system, one of two systems associated with the phenomenon of hybrid dysgenesis (4). An additional class of transposable elements called *I* factors (5) are associated with a second system of hybrid dysgenesis, the inducer–reactive (*I–R*) system. Both genetic and molecular evidence indicates that the two hybrid dysgenesis systems are essentially independent of one another (3, 6).

Although the emphasis in much of previous research on transposable element families has been on their molecular properties, we are now beginning to consider the population dynamics and possible evolutionary significance of such dispersed sequences. The phenomenon of hybrid dysgenesis, which appears in hybrids between certain mutually interacting strains, offers particular promise for the study of population and evolutionary problems. Hybrid dysgenesis is manifested in a number of correlated aberrant genetic traits such as high frequencies of sterility, male recombination, and mutation (4, 5). The specific sets of traits associated with each of the two hybrid dysgenesis systems are similar in general properties, but they differ in detailed characteristics. The genetic events leading to these traits are almost always restricted to the germ line and to one of the two classes of hybrids produced by reciprocal crosses between different interacting strains.

Every strain of *D. melanogaster* may be characterized with respect to its potential within each of the two systems of hybrid dysgenesis. In the *P–M* system, three classes of strains, *P*, *Q*, and *M*, have been described on the basis of their functional properties (6, 7). Hybrids between *P* strain males and *M* strain females show significant frequencies of gonadal sterility (8–10). *Q* strains do not show gonadal sterility in any strain combinations but produce male recombination and other dysgenic traits in crosses with *M*-strain females. *Q* strains are hypothesized to constitute a subset of the *P* family of elements that have lost their functional potential for sterility but have retained other *P*-element functions. All *P* and *Q* strains so far examined carry 30–50 copies of the *P* family of transposable elements (3). With one exception, all long-established laboratory *M* strains examined were completely lacking homology with the *P*-element family (3). *P* elements tend to be destabilized in the maternally derived cytoplasmic background of an *M* strain (*M* cytotype) but are stable within a *P* strain (*P* cytotype) (2, 11).

In the *I–R* system, two main classes of strains have been distinguished, inducer (*I*) strains and reactive (*R*) strains. A third group, referred to as neutral (*N*) strains, appears to be a weak subset of the *R* class (5). Hybrids between *I*-strain males and *R*-strain females show *SF* sterility, a type of sterility resulting from embryo lethality (12) and quite distinct from the gonadal sterility associated with the *P–M* system (8–10). Unlike the *P–M* system, dysgenic traits associated with the *I–R* system are always restricted to female hybrids.

The purpose of this paper is to report the results of an extensive survey of *D. melanogaster* strains with respect to their dysgenic potential. Possible implications of the observed historical and geographical patterns of distribution for the evolution of hybrid dysgenesis are discussed.

MATERIALS AND METHODS

Collection of Strains. The majority of the *D. melanogaster* strains used in the survey were obtained with the cooperation of scientists working with *Drosophila* throughout the world. These strains had diverse temporal and geographical origins in natural populations; the range of laboratory age was from 0 to 60 years, and most major geographical areas of the world were-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

represented. A detailed listing of the individual strains and the test results on which this survey was based will be published elsewhere.

Method of Strain Classification. Each strain of unknown characteristics was tested for its potential with respect to both hybrid dysgenesis systems. This was accomplished by mass matings in four separate crosses to standard *I*, *R*, *P*, and *M* reference strains. *I* (inducer) strains were defined as those which produced <10% eggs that failed to hatch (12) in crosses with males of the standard *I* strain, Luminy, but produced at least 10% eggs that failed to hatch from crosses with females of seF_8 or Cockaponsett Forest standard *R* strains. *R* (reactive) strains were defined as those which had no inducer activity as defined above but which produced at least 10% eggs that failed to hatch in the F_1 female progeny from crosses with Luminy males. *N* (neutral) strains produced <10% eggs that failed to hatch in *both* test crosses described above. *N* strains are considered to be a subset of *R* strains. *P* and *Q* strains were defined as those that produced, respectively, >10% and <10% rudimentary ovaries (8, 10) in samples of the F_1 female progeny from crosses with females of Canton-S, a standard *M* strain, but no substantial frequencies of rudimentary ovaries in F_1 females from crosses with males from Harwich, a standard *P* strain. *M* strains were defined as those which had no *P* strain activity when crossed with Canton-S females but which produced >10% rudimentary ovaries in the F_1 female progeny from crosses with Harwich males. Frequencies of the two types of reduced fertility were estimated by using developmental temperatures of 20°C and 29°C, respectively. Further details of the reference strains used and method of classification are provided in ref. 6.

The survey was started in 1977. Strains collected from the wild before that date had thus been maintained under laboratory conditions for variable periods of time prior to classification. Strains originating from the wild after that date were tested and classified as soon as practical after entering our laboratory.

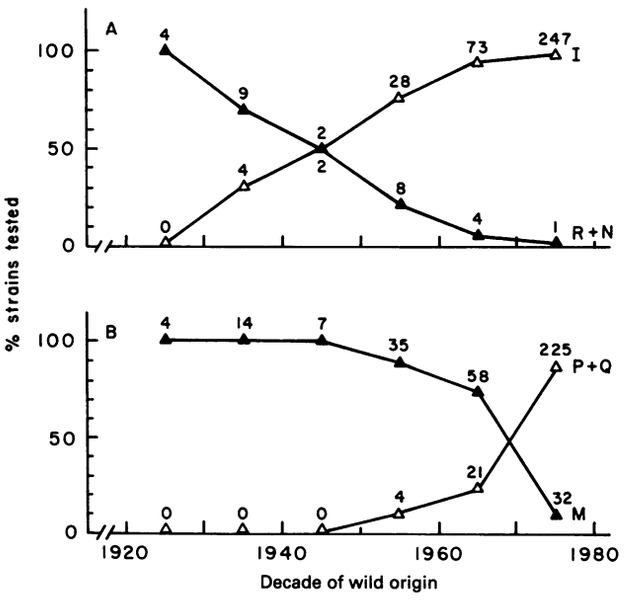

FIG. 1. Temporal distribution of strains collected worldwide according to their potential for the two systems of hybrid dysgenesis. (*A*) The *I–R* system in which the frequency of unhatched eggs (*SF* sterility) was used for strain characterization. (*B*) The *P–M* system in which the frequency of rudimentary ovaries (*GD* sterility) was used for strain characterization. The number of strains tested is indicated above each point on the graphs.

RESULTS

The tested strains were first analyzed as a single group for temporal trends, irrespective of their geographical origin, and later divided for further analysis into three major geographical categories: (*i*) continental America; (*ii*) Europe, Africa, and the Middle East; and (*iii*) Australia and the Far East.

***I–R* System.** The frequencies of *I* and *R* strains according to their time of origin in the wild are shown in Fig. 1*A*. The *I* type was not represented among the oldest laboratory strains and first appeared in strains originating in the wild about 1930. In strains originating during the period 1930–60, *R* and *N* types rapidly decreased, and the *I* type proportionately increased in frequency with decreasing laboratory age. Furthermore, no strain originating in the wild during the last decade was clearly identified as being of the *R* type. These results are consistent with those of previous studies of the distribution of *I* and *R* strains in recently derived natural populations (13).

***P–M* System.** Fig. 1*B* shows the frequency of *P* and *M* strains from all over the world together according to their decade of origin in the wild. Analogous to the pattern for the *I–R* system, there is a striking association of the frequency of *P* and *Q* strains with laboratory age. However, comparison of the distributions in Fig. 1 *A* and *B* provides two notable differences. The earliest collection to display *P* or *Q* activity in recent tests was made about 1950. This was approximately 20 years later than the date of the earliest collection of a strain showing *I* factor activity. Also, in contrast to *R* strains, *M* strains exist in present-day natural populations.

Temporal changes in the frequencies of *P*, *Q*, and *M* strains

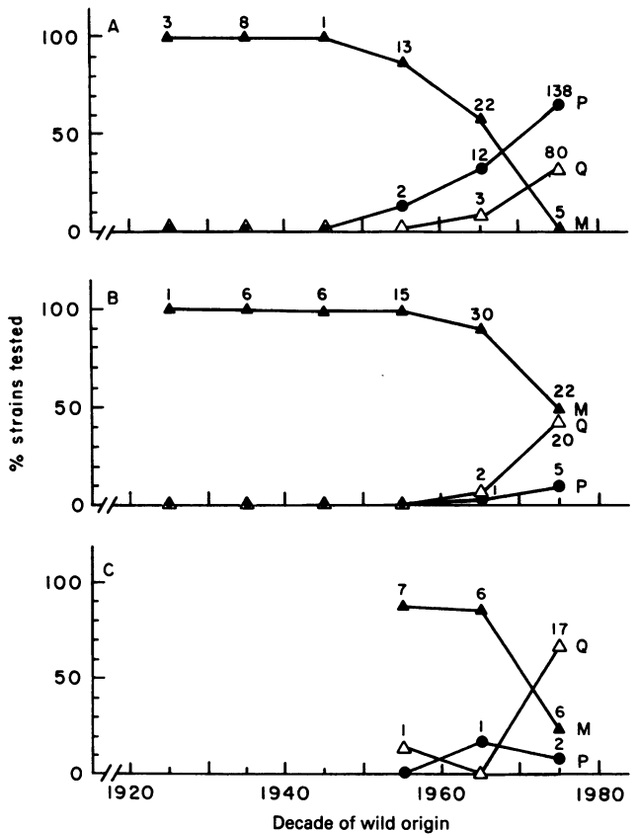

FIG. 2. Temporal distribution of strains with respect to their potential for the *P–M* system. (*A*) North and South America. (*B*) Europe, Africa, and the Middle East. (*C*) Australia and the Far East. *P*, *Q*, and *M* strains were defined as in Fig. 1. The number of strains characterized is indicated above each point on the graphs.

were examined separately for each of the three major geographical areas of the world (Fig. 2 *A–C*). In contrast to the *I–R* system, there are marked geographical differences in temporal trends for *P*, *Q*, and *M* strains, particularly between the American continent and the rest of the world. In the Americas, *M* strains are presently very rare in natural populations, and the frequency of the *P* type is consistently higher than that of the *Q* type in strains originating during the last 30 years. The distributions of strains from other geographical areas differed from the American distribution in that *P* strains were found at only low frequencies and *Q* strains showed a progressive increase in frequency during this same period. In addition, *M* strains, which were rare in recent American samples, were present in substantial frequencies in Europe, Africa, and Australia.

All the data reported in Figs. 1 and 2 were based on the analysis of wild-type strains. Similar analyses of a sample of long-established laboratory marker and balancer strains gave the following results. Of 78 strains examined, 54 (70%) were reactive or neutral and 24 (30%) were inducer. Only one of the same 78 strains was of the *P* type; the rest were all of the *M* type. Although the precise dates of synthesis or natural origin of these strains are unknown, the majority were in existence prior to 1950. Thus, the frequencies of *I* and *P* types within these laboratory marker strains is in good general agreement with those of wild-type strains originating during the 1920–1950 period (See Figs. 1 and 2).

Quantitative Changes in *P* and *M* Strains. In addition to qualitative temporal changes in strain distribution, there is evidence that quantitative variation within strain types exists. Fig. 3 *A* and *B* shows the distributions of mean sterility frequencies within samples of *M* strains that were collected from the wild before and after 1970, respectively. The strains originating before 1970 uniformly possessed strong *M*-strain properties at the time of testing, whereas those originating more recently from nature had, on the average, weaker and more variable *M* properties. Detailed analysis of the distribution of sterility frequencies among the F_1 hybrid dysgenic progeny of individual females from a number of weak *M* strains suggested an occasional clearcut polymorphism for *M* and *P* cytotype but, more often, continuous quantitative variability in cytotype.

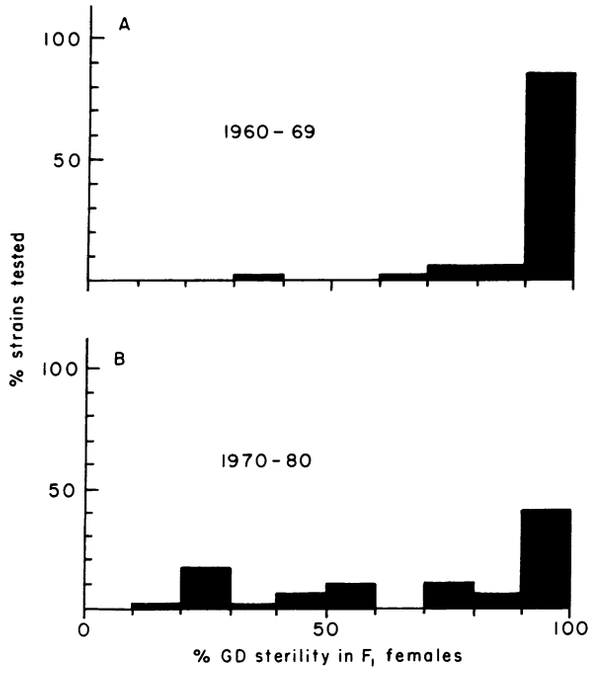

FIG. 3. Quantitative variation among different *M* strains according to their decade of origin in the wild. (*A*) *M* strains collected between 1960 and 1969 (56 tested). (*B*) *M* strains collected between 1970 and 1980 (31 tested). The frequency of F_1 rudimentary ovaries (*GD* sterility) was determined by mating females of each strain with males of Harwich, a standard *P* strain.

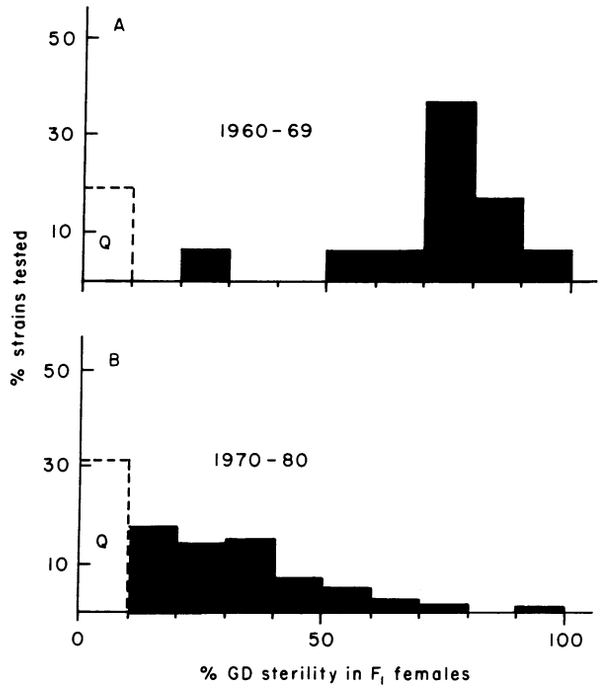

FIG. 4. Quantitative variation among different American *P* strains according to their decade of origin in the wild. (*A*) *P* and *Q* strains collected between 1960 and 1969 (16 tested). (*B*) *P* and *Q* strains collected between 1970 and 1980 (246 tested). The frequency of F_1 rudimentary ovaries (*GD* sterility) was determined by mating males of each strain with females of Canton-S, a standard *M* strain. The frequency of *Q* strains is shown by dotted lines at the left of the *P*-strain histogram.

A similar comparison of the distribution of two age groups of *P* strains collected in the Americas (Fig. 4 *A* and *B*) indicates that *P*-factor activity of those originating prior to 1970 was, on average, considerably stronger at the time of testing than that of those originating in the wild during the last decade.

DISCUSSION

Two main types of hypotheses have been proposed previously to explain the relationship between strain distribution and laboratory age—the recent-loss hypothesis and the rapid-invasion hypothesis. The recent-loss hypothesis contends that the observed distribution patterns can be accounted for by changes that have occurred in the laboratory. It proposes that *I* and *P* elements always have been present in substantial frequencies in natural populations, at least within recent historical times, and that the observed changes in the distribution of strains over the last few decades are the result of loss of *I* and *P* elements because of differences between natural and laboratory conditions. One possible mechanism for this type of change is the stochastic loss of elements in small populations (14). The second hypothesis, the rapid-invasion hypothesis (3, 6), proposes that, until recently, *I* and *P* elements have been absent or present only in low frequencies in natural populations of *D. melanogaster*. Further, these *I*- and *P*-element families are considered to have recently spread and rapidly increased in frequency in the wild. The mechanism of rapid invasion is postulated to be high efficiency of replicative transposition with or without positive or negative selection. Although these two main hypotheses

for distribution changes are inherently distinct, the rapid-invasion hypothesis and the mechanism of stochastic loss are not mutually exclusive. The rapid-invasion hypothesis does not rule out a limited frequency of stochastic loss of *I* and *P* elements in either small isolated natural populations or in laboratory populations. It states, however, that such stochastic losses of entire elements cannot account for the major distributional trends as shown in Figs. 1 and 2.

The results of the present survey may be explained by either major hypothesis but in two quite different ways. The changes in frequency of *I* and *P* strains associated with their time of collection from nature may be interpreted in terms of loss of *I* and *P* elements in laboratory environments, *P* elements being lost, on average, more quickly than *I* elements. Alternatively, they may be interpreted as indicating a rapid recent increase in natural populations in the frequencies of strains carrying the *I* and *P* families of transposable elements. According to this interpretation, the periods of rapid increase for the two families are overlapping but not completely synchronous.

There are a number of difficulties in explaining the present results by complete loss of *I* and *P* transposable elements in laboratory cultures. By assuming that the range in *P*-element copy number in *P* and *Q* strains is 30–50 and that most laboratory *M* strains are completely lacking in *P* elements, as suggested by recent results (3), then the laboratory evolution hypothesis would require that each *M* strain lose every one of the 30–50 copies of the *P* element during a short period in the laboratory. A number of strains have been monitored directly after collection from the wild at regular intervals during the last 5–10 years. It may be argued that insufficient time has elapsed to detect changes with respect to the *I–R* system in strains originating in the last decade. However, more rapid loss of *P* than *I* elements would be expected. Although a change from *P* to *M* has been observed in one stock (15), evidence has not been forthcoming that changes in laboratory stocks have occurred of either a magnitude or consistency of direction required to support the recent-loss model. On the other hand, the presently reported observation that *P*-factor activity is, on average, stronger in older than in younger laboratory strains is consistent with this model, given that sufficient generations have elapsed.

There are three main conceptual difficulties with the rapid-invasion hypothesis. First, the low Darwinian fitness of dysgenic hybrids might be considered to be a sufficient impediment to the rapid global spread of the *I* and *P* transposable elements. However, transposition frequency may be so high as to overcome this fitness reduction, particularly under permissive temperature conditions (6). By using a population genetics model, it has been shown (16) that selection against heterozygotes carrying a transposable element can be offset by a relatively high rate of transposition. Further, simulation studies by M. K. Uyenoyama and M. Nei (personal communication) have demonstrated that a transposable element, such as the *P* factor, can exist in a population for a long time at low frequencies, which increase only very slowly (and therefore may not be detected by limited sampling); above a critical frequency, the element may increase rapidly and sweep to fixation in a short time period. The results of a laboratory study (17) provide strong evidence that changes from the *M* to *P* type can occur rapidly, at least in small populations and under permissive temperature conditions, as proposed by the rapid-invasion hypothesis. The second difficulty is in relation to the likelihood of global spread of elements in a short time period. However, *D. melanogaster* is a species which is commensal with man, and global migration is greatly facilitated by intercontinental export of fruits. Both the geographic variability related to the *P–M* system and the presence of *M* strains in current natural populations are consistent with the rapid-invasion hypothesis.

The third difficulty is that, if new transposable element systems invade and accumulate in the genome at a rate of about two per century as suggested by the rapid-invasion hypothesis, then the number of elements expected to accumulate per genome over evolutionary time is unreasonably large. However, we do not know whether the putative recent invasion of natural populations by hybrid dysgenesis determinants is indicative of an average rate. Recent global environmental changes may have accelerated substantially the rate of invasion. Alternatively, the rate observed during a single time period may vary considerably from the average rate, merely by chance.

An understanding of the evolution of hybrid dysgenesis determinants requires the synthesis of information from both the population and molecular levels. Recent results on the molecular basis of *P–M* hybrid dysgenesis suggest that strains functionally characterized as being of the *M* type are not all identical at the level of DNA sequences. An exceptional strain has been observed (3) that was functionally characterized as *M* but nevertheless had some structural homology with a cloned *P* element. It has been hypothesized (3) that this strain carried a deficient *P* element that had lost some if not all of its *P*-factor functions. It was further hypothesized that, in contrast to most long-established laboratory *M* strains, which completely lack any homology to the *P*-element family, a large fraction of *M* strains recently collected from nature might be of this type (i.e., functionally *M* but having some *P*-element sequences). This prediction has been fulfilled in a preliminary study of a small sample of *M* strains from natural populations; four relatively weak *M* strains from Australia and Russia were similar to the original exceptional *M* strain in having reiterated copies of the *P* element (P. M. Bingham, personal communication).

From this result and other observations indicating the lability of the *P*-element family in *M* cytotype (2, 18), it is suggested that the quantitative variation observed in *M*-strain phenotype (Fig. 3) may reflect important underlying qualitative variation at the molecular level. In contrast to *M* strains lacking *P* elements, which have the potential for the maximum of 100% sterility in dysgenic hybrid progeny, those with a potential for less than the maximum sterility may carry a variable number of *P* elements that are defective because of deletion of internal element sequences and that specify corresponding variable degrees of *M* cytotype. Thus, the decrease in mean "strength" of *M* strains during the last decade (Fig. 3*B*) may be explained by an increase in frequency of *M* strains carrying defective *P* elements relative to those *M* strains completely lacking any copies of the *P*-element family.

The survey results provided evidence for variation of strain distribution patterns within the *P–M* system related to geographical origin in present-day natural populations. One of the major implications of this geographical variation is that dysgenic traits are expected to occur only exceptionally in matings within or between present-day American natural populations because of the observed low frequency of *M* cytotype. This suggests that the type of mutator activity reported previously (19), if confirmed, may have had a different origin. Conversely, hybrid dysgenesis may be more common in subpopulation crosses in other parts of the world where polymorphisms for *P* and *M* and for *Q* and *M* types exist.

On the basis of the present observations, taken together with other recent findings (2, 3, 20), an extended version of the rapid invasion hypothesis is proposed. Prior to about 1930, it is suggested that the *I* and *P* families of transposable elements were absent or present only in low frequencies in all or almost all natural populations, which were thus essentially of the *RM* type. About 50 years ago, *I* factors began to rapidly invade natural

populations. The invasion was completed by about 1960 after which time *I* strains have been essentially ubiquitous in the wild. A similar invasion by the *P* transposable element family is postulated to have begun about 30 years ago. Further, the originally invading *P* element family possessed all or most known functions of *P* elements, including the ability to transpose, to achieve cytotype conversion, and to induce sterility and other dysgenic traits. However, individual members of the *P*-element family are known to be heterogeneous in size (2, 3) and function (18); and it is suggested that loss of some or all of the original functions of *P* elements (possibly by internal deletion of element sequences) in *M* cytotype may have frequently accompanied the invasion in some geographic areas. This could have led to the evolution in some areas of *Q* strains possessing all functions except sterility and in other areas of modified or weak *M* strains essentially lacking *P* functions but having residual homology to the *P* element. Thus, the observed present-day polymorphism for *P*, *Q*, and modified *M* strains might be explained by the complete invasion of most natural populations by the *P*-element family. However, the size, structure, and function of individual elements would be expected to vary widely because of internal deletion of element sequences.

A mechanism by which successive invasions by transposable element families, such as the hybrid dysgenesis determinants, could lead to partial or complete reproductive isolation has been proposed (3). After geographic isolation, subpopulations of a single species are postulated to accumulate different transposable element families, each having the potential for partial hybrid sterility in interpopulation hybrids. On regaining sympatry, the combined effect of hybrid dysgenesis resulting from the destabilization of multiple transposable element families might, given sufficient divergence time, result in complete reproductive isolation. Such a process is considered to be typical of a "genomic disease mechanism," one of several proposed molecular biological mechanisms of speciation (21). Populations having lost or failed to acquire a transposable element family are considered to lack some property of "immunity" to the potential disruptive action of that family. Germ-line abnormalities would occur in hybrids between immune and susceptible populations, as defined by the presence or absence of the element, leading to postzygotic reproductive isolation. One of the current outstanding questions concerns the frequency with which such mechanisms exist in *Drosophila* and other taxa.

I thank J. B. Novy and T. Frydryk for technical assistance; Drs. P. Kareiva, W. R. Engels, and M. R. Rose for comments on the manuscript; and the many *Drosophila* workers who generously supplied their strains. This research was supported by the National Science Foundation (DEB-76-82630) and the National Institute of General Medical Sciences (GM-25399).

1. Spradling, A. C. & Rubin, G. M. (1981) *Annu. Rev. Genet.* **15,** 219–264.
2. Rubin, G. M., Kidwell, M. G. & Bingham, P. M. (1982) *Cell* **29,** 987–994.
3. Bingham, P. M., Kidwell, M. G. & Rubin, G. M. (1982) *Cell* **29,** 995–1004.
4. Kidwell, M. G., Kidwell, J. F. & Sved, J. A. (1977) *Genetics* **86,** 813–833.
5. Bregliano, J. C. & Kidwell, M. G. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. A. (Academic, New York), pp. 363–410.
6. Kidwell, M. G. (1979) *Genet. Res.* **33,** 205–217.
7. Engels, W. R. & Preston, C. R. (1981) *Drosoph. Inf. Serv.* **56,** 35–37.
8. Engels, W. R. & Preston, C. R. (1979) *Genetics* **92,** 161–174.
9. Kidwell, M. G. & Novy, J. B. (1979) *Genetics* **92,** 1127–1140.
10. Schaefer, R. E., Kidwell, M. G. & Fausto-Sterling, A. (1979) *Genetics* **92,** 1141–1152.
11. Engels, W. R. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4011–4015.
12. Picard, G., Lavige, J. M., Bucheton, A. & Bregliano, J. C. (1977) *Biol. Cell.* **29,** 89–98.
13. Bucheton, A., Lavige, J. M., Picard, G. & L'Héritier, P. (1976) *Heredity* **36,** 305–314.
14. Engels, W. R. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45,** 561–565.
15. Engels, W. R. & Preston, C. R. (1980) *Genetics* **95,** 111–128.
16. Hickey, D. A. (1982) *Genetics* **101,** 519–531.
17. Kidwell, M. G., Novy, J. B. & Feeley, S. M. (1981) *J. Hered.* **72,** 32–38.
18. Rubin, G. M. & Spradling, A. C. (1982) *Science* **218,** 348–353.
19. Thompson, J. N., Jr., & Woodruff, R. C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1059–1062.
20. Anxolabéhère, D., Nouaud, D. & Periquet, G. (1982) *C. R. Hedb. Seances Acad. Sci. Ser. III* **294,** 913–918.
21. Rose, M. R. & Doolittle, W. F. (1983) *Science*, in press.