A theoretical assessment of recombinational speciation

EUGENE M. McCARTHY, MARJORIE A. ASMUSSEN* & WYATT W. ANDERSON

Department of Genetics, University of Georgia, Athens, GA 30602, U.S.A.

Using a computer simulation, we have examined the dynamics of recombinational speciation, a potentially rapid mode of evolution dependent on chromosomal reassortment in populations of partially sterile interspecific hybrids. We describe how various parameters affect the time required for a new recombinant species to become established within the setting of a spatially structured hybrid zone. Our results indicate that recombinational speciation is most likely to occur where (1) the hybrid zone interface is long, (2) the organisms involved are predominantly selfing, (3) the hybrids are relatively fertile, and (4) the number of differences in chromosomal structure between the parental species is small. The speciation dynamics are characterized by long-term stasis followed by an abrupt transition to a new reproductively isolated type. The results are largely the same whether the nascent recombinant species is favoured by a fertility or a viability advantage. Recombinational speciation, like polyploidy, appears to be a feasible mechanism for sympatric speciation in plants.

Keywords: chromosome rearrangements, hybridization, plant evolution, speciation, sympatric speciation.
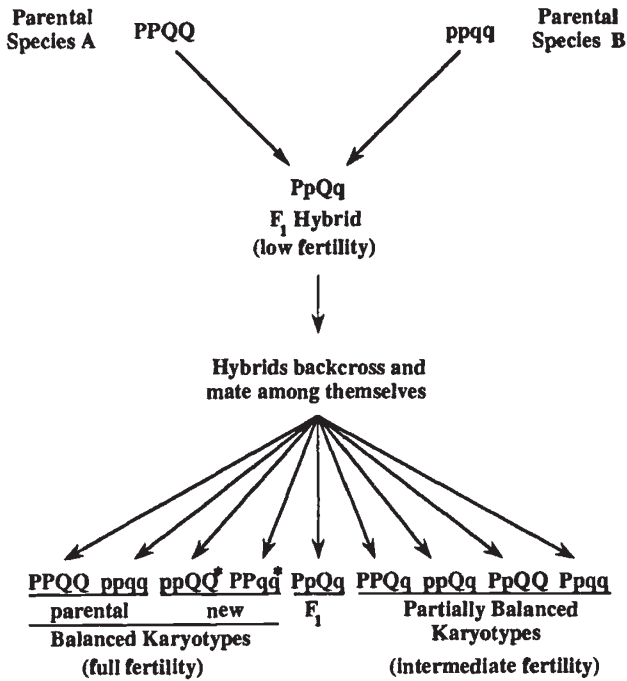
Introduction

Recombinational speciation is a process by which a new, reproductively isolated species can be derived, without change in ploidy level, from two or more parental species via hybridization. It has been formally defined by Grant (1985) as 'the formation, in the progeny of a chromosomally sterile species hybrid, of a new, structurally homozygous recombination type that is fertile within the line but isolated from other lines and from the parental species by a chromosomal sterility barrier'. When these new types are at the same ploidy level as the parental species they are termed 'homoploid derivatives'. This process has been used experimentally to produce new lines of plants descended from highly sterile hybrids (Stebbins, 1957, 1958; Grant, 1966a,b). The specifics of the process and its significance as a mode of evolution in plants have been discussed in a series of papers and books by Grant (1958, 1963, 1966b, 1971, 1985).

For recombinational speciation to occur, (1) the parental species must differ by at least two independent chromosome rearrangements (translocations or inversions), and (2) crossing-over in their structurally heterozygous progeny must reduce the number of

viable gametes. This reduction results from the duplication or deletion of large chromosomal segments and the production of acentric or dicentric chromosomes during meiosis. As an example (following Grant, 1985), suppose two parental species differ by two such structural rearrangements, P, p, and Q, q. Their F_1 hybrid will then be the double structural heterozygote PpQq. Six additional types can be produced in the F_2 or later generations, as shown in Fig. 1. Four of these hybrids (PPQq, ppQq, PpQQ, Ppqq) will be single heterozygotes and presumably sterile to some degree. Two, however, will be homozygous (ppQQ and PPqq) and should be fully fertile. These two homoploid derivatives are separated by a partial sterility barrier from both parental species, and from each other as well, because matings between any two of these types will produce heterozygotes of reduced fertility. The reproductive barrier is enhanced when larger numbers of independent chromosomal differences are involved. Inasmuch as they are both fertile *inter se* and at least partially isolated from other homozygous lines, such derivatives have the potential to become new 'recombinant' species if they ever become numerous enough to form a stable population.

A theoretical analysis of recombinational speciation would help in making judgements about the overall feasibility of this evolutionary mechanism, as well as

Parental
Species A   PPQQ   ppqq   Parental
                                                        Species B

PpQq
F₁ Hybrid
(low fertility)

Hybrids backcross and
mate among themselves

PPQQ ppqq ppQQ* PPqq* PpQq PPQq ppQq PpQQ Ppqq

<u>parental</u>   <u>new</u>   F₁   Partially Balanced
<u>Balanced Karyotypes</u>   Karyotypes
(full fertility)   (intermediate fertility)

* New partially reproductively isolated diploid species.

Fig. 1 The generation of reproductively isolated homoploid derivatives by recombinational speciation. The simplest case is shown, where the two parental species differ by only two independent chromosome rearrangements (P/p and Q/q).

the specific biological conditions promoting it. Grant (1971) originally postulated that the mating system would play an all-important role in the process, with homoploid derivatives emerging most rapidly and having maximal chances of surviving as a distinct population under inbreeding. The presumed requirement of a significant level of self-fertilization has been called into question, however, by recent studies indicating that two obligately outcrossing species had their origins in recombinational speciation (Gallez & Gottlieb, 1982; Rieseberg *et al.*, 1990). Furthermore, it is clear that a number of other factors are likely to play an important role, such as the degree of sterility in the F₁ hybrids and backcrosses, fitnesses of the new structural homozygotes relative to the two parental species, number of chromosomal differences distinguishing the parental types, and size of the hybrid zone. We have designed a simulation to evaluate the effect of these factors on the dynamics of recombinational speciation, and this paper presents our results.

Description of the model

The model is both spatial and stochastic. It simulates breeding in a hybrid zone between two monoecious

diploid plant species with discrete, nonoverlapping generations and no seed dormancy. Mating can occur by selfing or outcrossing. To test its effect, the selfing rate (s) was varied over the interval $[0,1)$. Because the model is spatial, matings are carried out on an individual basis. When outcrossing occurs, near neighbours are more likely to mate. Essential steps in the reproductive process are all stochastically determined: the fitness of individuals in terms of the number of progeny they produce, the choice between selfing and outcrossing, the selection of mating pairs, and the positioning of offspring. Every effort has been made to make the probability distributions associated with each of these steps empirically reasonable.

The modelled population is composed of two karyotypically homogeneous parental species hybridizing in the region where they interface. They are assumed to differ by two or more independently segregating chromosome rearrangements. The exact number of differences can vary from one simulation to another since this number (m) is one of the variables under consideration. The position of individuals within the population is specified by means of an indexed grid defining square cells superimposed over the hybrid zone (see Fig. 2). In all runs of the simulation the grid was 58 cells long, running from permanently pure species-A territory to permanently pure species-B territory. The *width* of the grid is actually the *length* of the interface between the two hybridizing species. This length (ℓ) was set to four cells for all runs, except those in which ℓ itself was the variable under consideration. Initially, five individuals were loaded into each cell with the interspecific boundary at the grid midline.

Each generation consists of three stages (see Fig. 3). First, each individual in the population is assigned a fitness. All individuals in the current generation then self or outcross to produce offspring. The number of offspring is then regulated, with the probability that an individual survives based on an environmental carrying capacity of five individuals per cell. Thus, if the number (c) of individuals in a given cell is five or less, all individuals in that cell survive to mate; otherwise, each survives with probability $5/c$.

Grant (1966a) created new, reproductively isolated lines by selecting for vigour and fertility. If recombinational speciation is to occur in a hybrid zone, a new, isolated line (new species) must emerge under the influence of *natural* selection. To implement such selection in the model, a particular homoploid derivative (e.g. PPqq in Fig. 1) is assumed to have a new, better-adapted combination of genes, resulting in a selective advantage over *all* other types (both hybrids and parentals). Such an assumption is reasonable inasmuch as derivatives more fit than either parental species are known from nature. For example, the

ℓ: Length of monitored portion of interface between species A & B

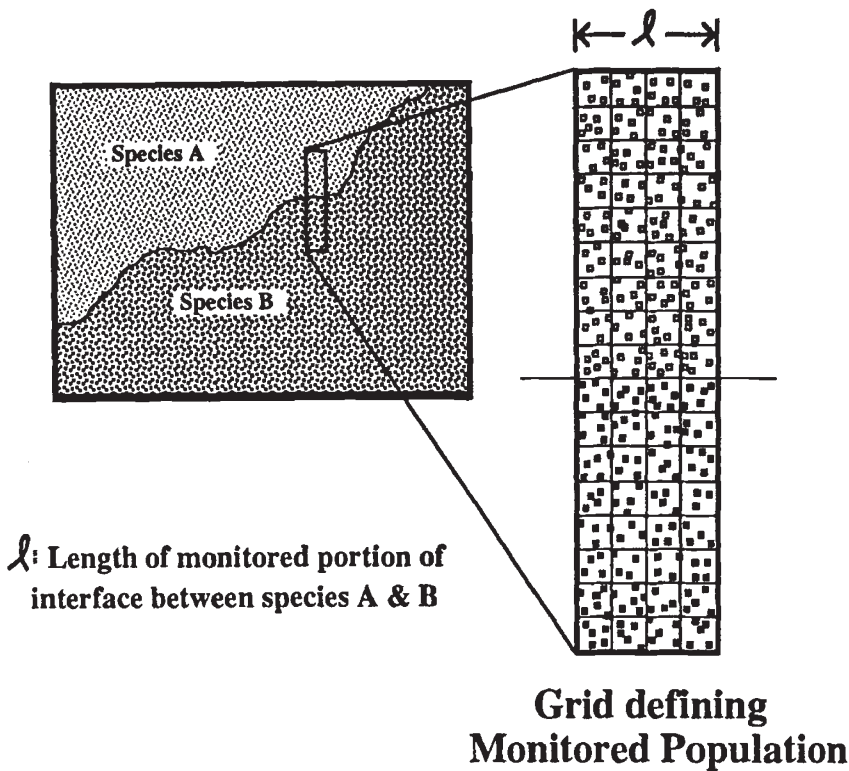Grid defining
Monitored Population

Fig. 2 Rectangular grid specifying positions of individuals within the population. The grid actually used in the simulation was much larger than that shown, containing 58 rows and a variable number (ℓ) of columns, as specified in the text.
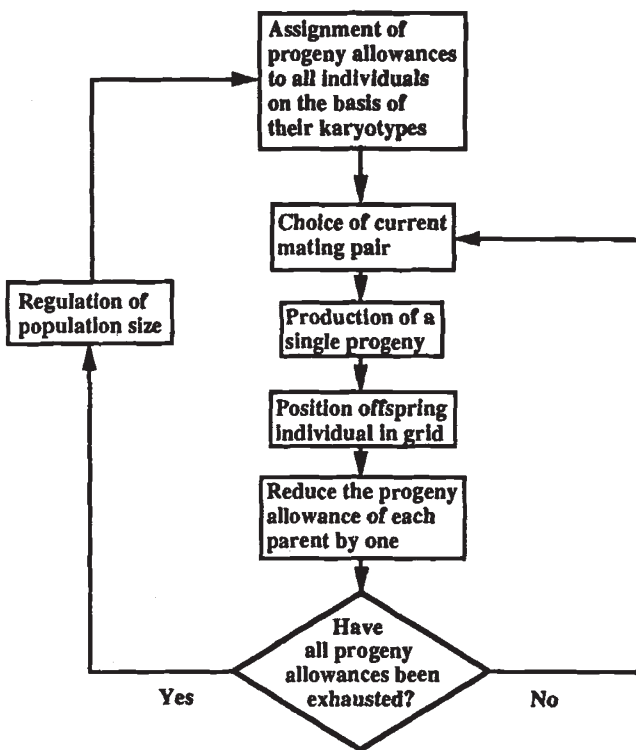

Yes   No

Fig. 3 Flowchart showing the major steps executed by the simulation each generation.

recombinant sunflower species *Helianthus paradoxus* is far more tolerant of saline conditions than either of the two species from which it is derived (Rieseberg *et al.*, 1990; Rieseberg, 1991, and Rieseberg, pers. comm.). We have investigated the effects of both fertility and viability selection in our model.

The assignment of fitnesses is carried out as follows: an expected fitness is associated with each of the possible karyotypes, but different individuals with the same karyotype can produce different numbers of offspring, since this number is assumed to be a Poisson random variable with parameter λ. Under fertility selection $\lambda = 4$ (yielding an intrinsic rate of increase of two before density regulation) for fully homozygous types with the exception of the selected homoploid derivative. For these individuals $\lambda = 4a$, where $a > 1$ is that type's fitness advantage relative to the parental species and other fully homozygous types. Individuals having $n \geq 1$ heterozygous chromosome pairs are assumed to produce progeny at a mean rate $\lambda = 4r^{n/m}$, where r is the fertility of the F_1 hybrid relative to its parental species and m is the number of independent chromosome rearrangements by which those species differ. This formulation is based on a simple rationale: the expected proportion of viable gametes produced by an individual should decrease multiplicatively with the

number of heterozygous chromosome pairs in that individual's genome.

The model is formulated somewhat differently when the process is instead driven by a viability advantage. Here the expected number of progeny λ is redefined to be $4ar^{n/m}$ for any individual with $n \geq 0$ heterozygous pairs. This formula increases the number of nonoptimal progeny initially generated by a factor a relative to the case of fertility selection, but after viability selection only a proportion $(1/a)$ of the nonoptimal types survive to mate. Under these conditions, the two selection scenarios are comparable in having the same relative strength of selection and roughly the same mean number of mating individuals each generation.

The probability distributions for the other critical stochastic events (i.e. choice of outcrossed mate and positions of offspring) are given in the Appendix. Complete details of the model are available from the authors on request.

Results

Our primary goal has been to determine (1) the conditions where a recombinant species is likely to become established and to spread; and (2) whether 'swamping' from backcrossing would be an insurmountable barrier to the stabilization of a new hybrid type. Answers to these questions would give us a better idea of how prevalent this mode of evolution might be in nature, as well as a set of characteristics one would expect to find in natural populations affected by recombinational speciation.

Our analysis focuses on the number of generations required for the new type to become established, given a particular set of fixed conditions, where the selected type is considered to be 'established' as soon as 100 or more such individuals are present in the population. This criterion was chosen because, in our simulations, the optimal type has always been observed to increase steadily to fixation after having reached this level. We have termed the number of generations required for the population to reach this point 'the number of generations to transition'. Because this number is the outcome of a random trial, it is a random variable and will be denoted by G. For each set of fixed variables, we report the average number of generations to transition, $\bar{G}$, taken over a number of different computer runs (24 except in those cases where variation between runs was minimal). Recombinational speciation is most likely to occur under those conditions where $\bar{G}$ is small. The illustrative examples presented below are all taken from the case where the process is driven by a fertility advantage. In general, when all other factors are held constant, the same qualitative results hold for the case of viability selection.

Effect of F_1 fertility

The first question investigated was whether selection for a new recombinant type can overcome low fertility in the hybrids and allow recombinational speciation to occur within the context of a spatially structured hybrid zone. We addressed this issue by examining the effect of r, the fertility of the F_1 hybrids relative to that of the parental species. The value of $\bar{G}$ corresponding to various values of r is shown in Fig. 4. In these runs the advantage of the selected type, a, was 1.25. The picture that emerges is that low fitness of heterozygotes does not actually prevent the new type from getting established. It merely slows the process. Only when heterozygotes are absolutely sterile is the process entirely blocked. Otherwise the average time required for transition increases steadily from the right-hand end of the graph, which represents no fertility disadvantage from heterozygous chromosome pairs, to the left-hand end, which represents absolute sterility in the F_1 hybrids. The increase in $\bar{G}$ appears to be asymptotic with the vertical axis, but does not appear to increase so rapidly that hybrids of low fertility would be unable to participate in the process. Even when r was as low as 0.018, the new species was established in just 386 generations, on average.
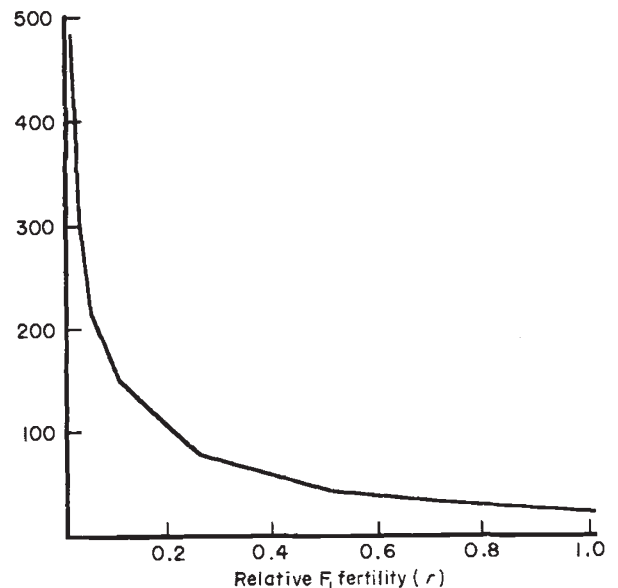

Fig. 4 Mean number of generations to transition, $\bar{G}$, plotted against the relative fertility of F_1 hybrids (r). Fixed variables settings: $\ell = 4$; $m = 2$; $a = 1.25$; $s = 0.85$.

In general, it should be noted that values of variables that increase $\overline{G}$ also increase the variance of G and the skewness of its distribution. In consequence, when $\overline{G}$ is large, most values of G fall below the mean and some, far below it.

Effect of selection, selfing and outcrossing

In all of the computer runs evaluating the effect of the fitness advantage α, the relative fertility (r) of F_1 hybrids was held constant at 0.1. Such an r-value is reasonable considering actual data on interspecific F_1 hybrids, where fertility varies widely depending upon the particular cross in question. The effect of the fitness advantage on $\overline{G}$ is much the same as that of r. When α goes up, $\overline{G}$ comes down.

The combined effects of selfing rate and selection are presented in Fig. 5. It is clear that the selfing rate has a strong effect on $\overline{G}$ for each selective value α examined. For any given value of α, $\overline{G}$ is small for a range of selfing rates, but increases rapidly for values of s below this range because the process is retarded by outcrossing. For values of s very close to 1, $\overline{G}$ increases asymptotically since no hybridization occurs in strict selfers (not shown in Fig. 5, but evident in the smaller scale of Figs 6 and 7). In general, decreasing α markedly increases the amount of selfing required for rapid transition, but when α is sufficiently large (e.g.
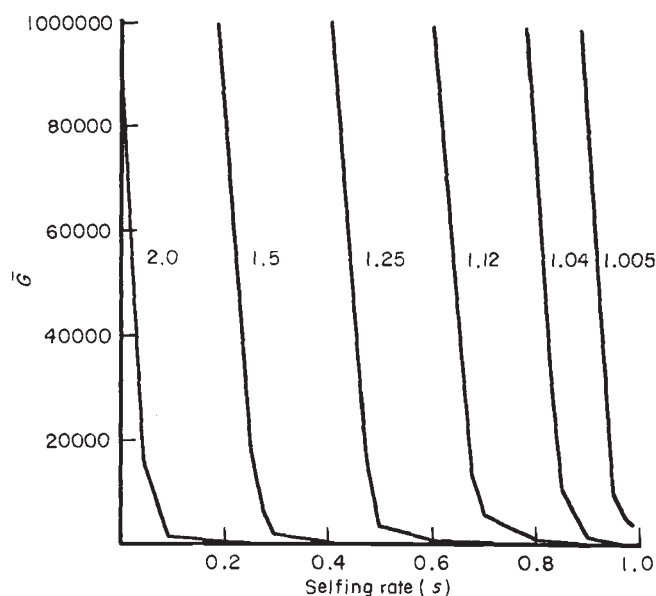
$\alpha = 2.0$), the process appears to be feasible even for obligate outcrossers.

It should be noted that the determination of a realistic range of values for α, the advantage of the selected type, is a problematic point. This variable is used in two ways in our analysis, as a fertility advantage, and as a viability advantage. Regarding the former, there appears to be very little empirical information on the fertility of homoploid derivatives relative to their parental species. Since it is difficult to say what a reasonable value for α might be, we have explored as broad a range of values as the speed of our computer allows (run times increase asymptotically as α approaches 1). In the case of viability selection, however, it is empirically apparent that a homoploid derivative can enjoy a significant advantage over its parental
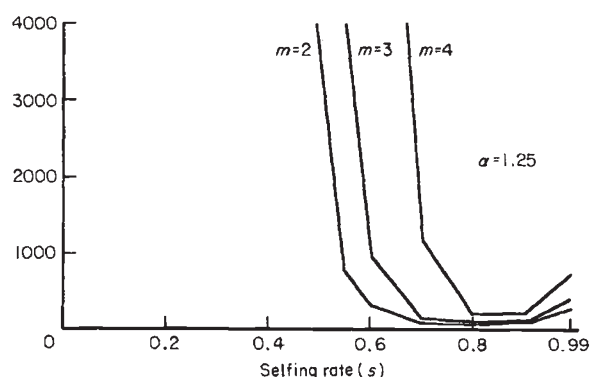

Fig. 6 Mean number of generations to transition, $\overline{G}$, plotted against the selfing rate (s) for various numbers of chromosomal differences (m). Other variables: $\ell = 4$; $\alpha = 1.25$; $r = 0.1$.
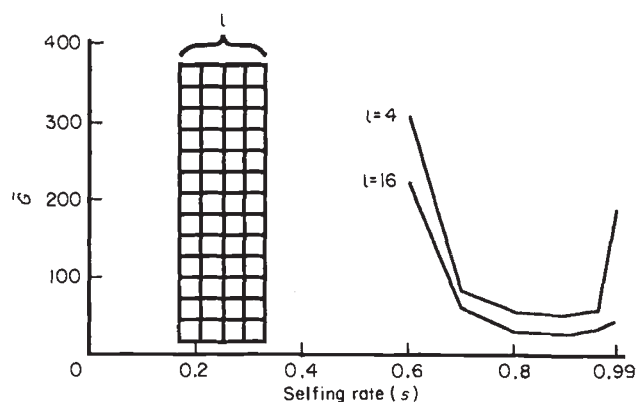

Fig. 5 Mean number of generations to transition, $\overline{G}$, plotted against the selfing rate (s) for a variety of selective advantages (α). In each case, the value of α is shown to the right of the line to which it corresponds. Other variables: $\ell = 4$; $m = 2$; $r = 0.1$.

Fig. 7 Effect of species interface length, ℓ. Mean generations to transition, $\overline{G}$, plotted against the selfing rate (s) for two values of ℓ. Other variables: $\alpha = 1.25$; $m = 2$; $r = 0.1$. The grid shown has far fewer cells than the one actually used (see text).

species. In the previously cited example of *H. paradox-us*, for instance, whose parental species can scarcely survive in the brackish marshes where their derivative thrives, it would not be unreasonable to assume a viability advantage of 10 to 25 per cent (Loren Rieseberg, pers. comm.). For pragmatic reasons, we have generally taken α to be 1.25 in the computer-intensive runs used to illustrate the effects of the other variables examined. In these other cases the qualitative results are generally the same for other α-values, except for the change in time-scale illustrated in Fig. 5.

Effect of number of chromosomal differences

$\bar{G}$ increases with the number of chromosomal differences, m (Fig. 6), with the magnitude of the increase approximately proportional to m. Our study thus indicates speciation is less likely to occur when the number of differences is large, an unexpected result which we will examine further in the Discussion section.

Effect of species interface length

By adjusting the width of the grid, we have investigated the effect of lengthening the interface between the two parental species. In general, the longer the interface, the shorter the time to transition. In Fig. 7, results are shown for two interface lengths, $\ell = 4$ and 16, where ℓ is given in number of cells. One can see that lengthening ℓ from 4 to 16 decreases $\bar{G}$, and, for most s-values, markedly so. Our overall results indicate that $\bar{G}$ is in an approximately inverse relation to the length of the zone interface. This result is, in fact, consistent with expectation: a transition is really the consequence of a particular, improbable series of hybrid matings. The longer

the interface, the more hybrid matings per generation, which in turn means that fewer generations should be required, on average, for the first string of matings to occur.

Stasis followed by sudden transition

Perhaps even more interesting than these quantitative results are the general dynamics of the process. A representative trajectory is shown in Fig. 8, which plots the number of individuals of the optimal type present in a single population over time. Examining the results of this run in detail, we find that during an initial period of stasis few or no individuals of the selected type are present in most generations, although their number does increase ephemerally at irregular intervals, sometimes even building up into short-lived blooms which persist for several generations before dying out. However, once the final transition begins, after about 3550 generations, a new pattern of steady, exponential increase emerges, during which the selected type spreads to fill the grid.

Discussion

We have analysed the effects of several variables on the process of recombinational speciation. In certain respects, the findings of our study are consistent with expectation, inasmuch as they indicate that the process is facilitated by higher fertility in the F_1 hybrids, a higher selective advantage for the optimal type, and a long hybrid zone interface. Our results are also broadly consistent with Grant's (1971) postulate that high selfing rates are conducive to recombinational speciation, although the process appears not to depend absolutely
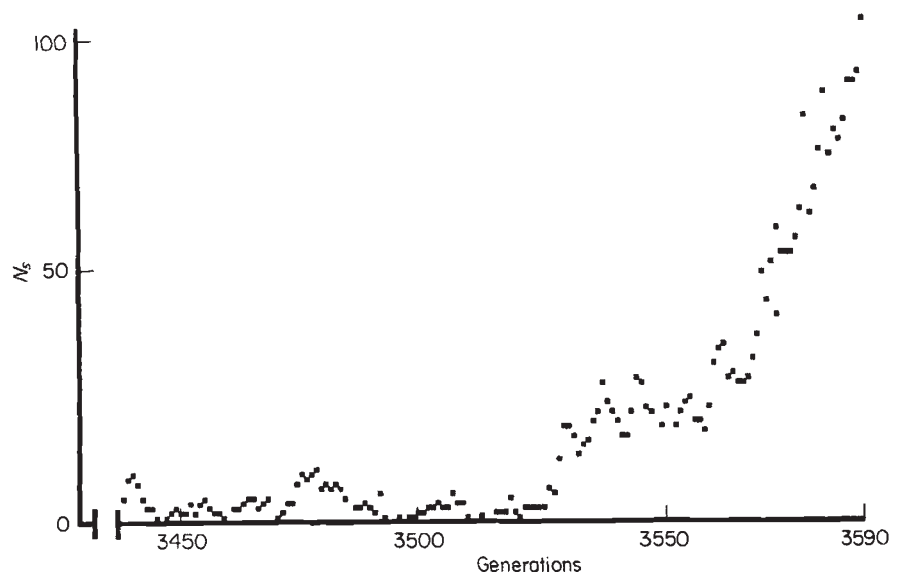
Fig. 8 Number of individuals, N_s, of the selected type present in a single population during each generation. The break in the horizontal axis indicates a long period during which the number of individuals of the selected type remained at or very near zero. Variables: $\ell = 4$; $m = 2$; $r = 0.1$; $\alpha = 1.25$; $s = 0.3$.

on selfing. Although these findings are perhaps not entirely surprising, they are enlightening by virtue of their more exact delineation of the circumstances under which recombinational speciation is likely to occur.

A more interesting and less anticipated result is the qualitative dynamics of the process, i.e. a pattern of long-term hybrid zone stasis followed by an abrupt transition where the selected type suddenly gains a foothold, mushrooms in number, and supplants the two previously existing species as it spreads to fixation. The phenomenon is reminiscent of the appearance and spread of a 'critical bubble' in a previously uniform, continuous population (Rouhani & Barton, 1987). A key factor underlying this change in dynamics seems to be the number of individuals of the optimal type present in the population. When their number is low, such individuals are unlikely to find mates of their own type and thus will not leave progeny of their own kind; they suffer the fate of Goldschmidt's hopeful monster. When they are more numerous, however, finding a suitable mate is not a problem and the selective advantage assigned to individuals of the optimal type guarantees their increase. Because the model is stochastic, the number of such individuals present fluctuates from one generation to another, sometimes becoming quite large just as a matter of chance. When large enough, these ephemeral increases have an apparent feedback effect. The increased number first improves the chances of finding a mate, and this improvement then further increases the number in the next generation. The result of this feedback cycle is an abrupt transition initiating the rapid, steady spread of the new type.

Another factor involved is the arrangement of the population in space. If individuals of the optimal type are clumped, they are more likely to mate among themselves than if they were evenly dispersed. A certain amount of clumping occurs simply because individuals of the selected type at first occur only where the two parental species interface. Both in the model and in nature, the size and the genetic purity of such clumps would be expected to vary from one location, or time, to another simply because the process is a random one. If any one of these clumps is sufficiently large to let the new species get established, a transition will occur. In general, then, it seems that transitions should be facilitated if the optimal type is concentrated both in space and time. Thus, a nonfluctuating, deterministic model that did not take into account the spatial distribution of individuals is not apt to generate the abrupt shifts in dynamics that we have described.

That the selected type is able to get established and spread under any circumstances is a surprising result, inasmuch as it represents a case of sympatric speciation. Although it is widely accepted that new plant species can arise sympatrically via polyploidy, sympatric speciation is considered impracticable in other contexts. Our findings suggest that sympatric speciation is also likely to occur without change in ploidy level within the setting of a hybrid zone, particularly when the participating organisms are capable of self-fertilization. In our model, the ability of the selected type to become established under such circumstances is presumably a consequence of its initial emergence in the form of concentrated clumps, which affords what one might term a 'microgeographical' isolation, in that clumping of similar individuals increases their propensity to mate with each other.

This clumping may account for what we considered to be a counterintuitive finding — that the process is facilitated when the number of chromosomal differences between the parental species (m) is *small*. One might think that a new hybrid type not isolated by a large number of chromosomal differences would be swamped out by backcrossing. We have found again and again, however, that the optimal type increases inexorably under the influence of selection, if it ever gets established at all, regardless of whether m is small or large. The number of differences only comes into play in the time required for the initial appearance of the new population. That the number of differences should play an important role *initially* is to be expected because the optimal type is only one of 3^m different karyotypes that can appear in the zone, making its chance emergence far less likely when m is large.

The increase in the probability of transition with the length of the hybrid zone suggests that recombinational speciation is likely to occur under a broader range of circumstances than a straightforward interpretation of our results would indicate. On a computer, time and memory constraints limit the monitored population to a size that would, in nature, be only a minute portion of even a small hybrid zone. In the case of a lengthy natural zone, where a very large number of hybrid matings occur each generation, one can imagine recombinational speciation taking place even when the characteristics (selfing rate, number of chromosomal differences, etc.) of the species participating are otherwise inimical to the process. The natural recombinant species identified by Rieseberg *et al.* (1990) and by Gallez & Gottlieb (1982), both derived from obligately outcrossing parental species, constitute cases in point, indicating that the requirements for recombinational speciation are probably less stringent for large, natural populations than one might conclude on the basis of our results alone. It should also be noted that the selection favouring the recombinant species was conservative in our simulations in that all other fully homozygous types were assumed to be equally fit. In actuality, there may be a gradation of selective differ-

ences, with fitness increasing with similarity to the optimal type, which would presumably facilitate that type's becoming established as a new species. Our findings, then, do not define preconditions for recombinational speciation in any absolute sense. Rather, they provide conservative guidelines specifying how likely it is that a given hybrid zone population will undergo the process.

Overall, our results suggest recombinational speciation is a workable mode of evolution, not only in the laboratory but also in nature. So long as the sterility of the hybrids is not absolute, the potential exists for the rapid reassortment of entire chromosomes into novel combinations whenever the new combination is advantageous. In this way, natural selection is able to exploit the rich genetic variability of the typical hybrid zone.

Acknowledgements

References

GALLEZ, G. P. AND GOTTLIEB, L. D. 1982. Genetic evidence for the hybrid origin of the diploid plant *Stephanomeria diegensis*. *Evolution*, **36**, 1158–1167.

GRANT, V. 1958. The regulation of recombination in plants. In: *Exchange of Genetic Material: Mechanisms and Consequences*, pp. 337–363. *Cold Spring Harb. Symp. Quant. Biol.*, vol. 23. Cold Spring Harbor, New York.

GRANT, V. 1963. *The Origin of Adaptations*. Columbia University Press, New York.

GRANT, V. 1966a. Selection for vigor and fertility in the progeny of a highly sterile species hybrid in *Gilia*. *Genetics*, **53**, 757–775.

GRANT, V. 1966b. The origin of a new species of *Gilia* in a hybridization experiment. *Genetics*, **54**, 1189–1199.

GRANT, V. 1971. *Plant Speciation*. Columbia University Press, New York.

GRANT, V. 1985. *The Evolutionary Process*. Columbia University Press, New York.

RIESEBERG, L. H. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *Am. J. Bot.*, **78**, 1218–1237.

RIESEBERG, L. H., CARTER, R. AND ZONA, S. 1990. Molecular tests of the hypothesized origin of two diploid species (Asteraceae). *Evolution*, **44**, 1498–1511.

ROUHANI, S. AND BARTON, N. 1987. Speciation and the 'shifting balance' in a continuous population. *Theor. Pop. Biol.*, **31**, 465–492.

STEBBINS, G. L. 1957. The hybrid origin of microspecies in the *Elymus glaucus* complex. *Cytologia*, Suppl. Vol., 336–340.

STEBBINS, G. L. 1958. The inviability, weakness, and sterility of interspecific hybrids. *Adv. Genet.*, **9**, 147–215.

Appendix

When an individual outcrosses, the probabilities of choosing its mate from a cell $0, 1, \ldots, 17$ cells away are given respectively by 0.4, 0.2, 0.1, 0.05, 0.025, 0.025, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.01, 0.005, 0.005. This represents a discrete approximation to a leptokurtic distribution. This distribution was chosen because a small fraction of a plant's pollen typically travels long distances even though most of it falls on near neighbours. The seed dispersal distribution is more compressed since seeds do not generally travel as far as pollen: seeds are assumed to fall at a distance of 0, 1, or 2 cells from the maternal plant with respective probabilities of 1/2, 1/3, and 1/6.