PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Use of Electronic Computation in the Study of Random Fluctuations in Rapidly Evolving Populations

J. L. Crosby

Phil. Trans. R. Soc. Lond. B 1960 **242**, 551-572
doi: 10.1098/rstb.1960.0002

THE USE OF ELECTRONIC COMPUTATION IN THE STUDY OF RANDOM FLUCTUATIONS IN RAPIDLY EVOLVING POPULATIONS

By J. L. CROSBY

Botany Department, Durham Colleges in the University of Durham

(*Communicated by D. G. Catcheside, F.R.S.—Received* 21 *July* 1959)

CONTENTS

Electronic computing techniques have been used to superimpose random fluctuations on calculations of the evolutionary sequence from heterostyly to homostyly in *Primula vulgaris* and on gene flow between populations. The effect of such fluctuations has been studied by considering isolated populations, and also groups of populations in an imaginary map. It is shown that the considerable irregularities in homostyle distribution which are found in natural populations can be accounted for by random fluctuations, and it has been possible to produce by electronic computation artificial maps which correspond very well with the real maps of natural homostyle distribution. An appendix outlines the computing methods used.

Introduction

The analysis by Sewall Wright of the effect of random fluctuations on gene frequency is well known. His attention was primarily directed to consideration of the relative frequencies of two alleles of a gene in populations which would be expected to tend towards an equilibrium with both alleles present. In his analyses, it was possible to provide a direct mathematical treatment, which would allow the expression of the results in a readily comprehensible way.

In the course of my work on homostyle primrose populations, it has become necessary to investigate theoretically the effect of random fluctuations in populations which are evolving rapidly at a variable rate and which involve three alleles (strictly, three allelic complexes) and four distinct genotypes. This is a much more difficult problem, and to an inexperienced mathematician it does not seem possible to provide a direct general solution which is expressible intelligibly and simply.

Vol. 242. B. 697. (Price 7*s*. 6*d*.) 68 [Published 17 March 1960

Evolution of homostyly in the primrose, *Primula vulgaris*

The problem has arisen in the following way (for a fuller account, see Crosby 1949). In certain areas in England, in addition to the usual and mainly outbreeding pin and thrum forms, primrose populations contain in widely varying proportions homostyle forms which are self-fertile and largely self-fertilizing. These homostyles are believed to be increasing in frequency at the expense of the normal forms, especially of thrums which are being completely eliminated. Evolution appears to be from populations containing only pins and thrums, to populations containing about 80 % of homostyles and 20 % of pins. This situation is illustrated in figure 1 which also describes the breeding system. The three phenotypes involve four genotypes: pins (ss), thrums (Ss), heterozygous homostyles ($s's$) and homozygous homostyles ($s's'$). The small amount of crossing between homostyles and thrums which is known to occur (Crosby 1959) is ignored in this account.

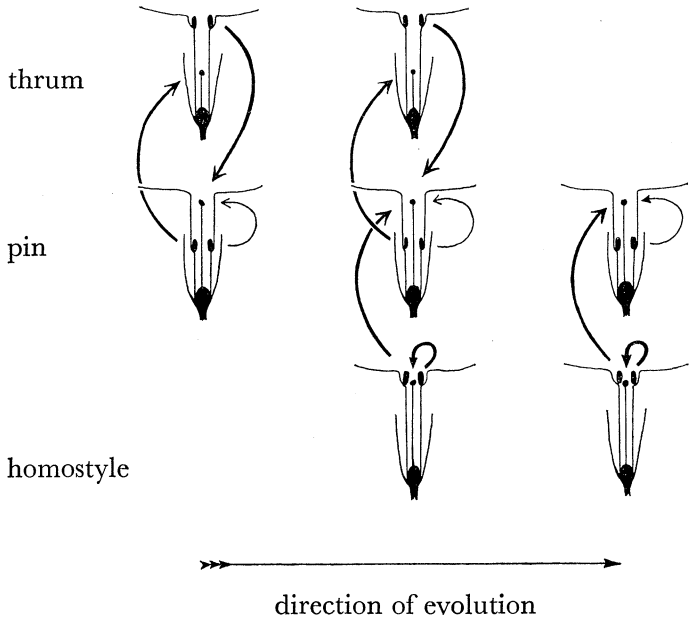


Figure 1. Diagrams of the three types of primrose and their breeding relations. The curved arrows show the effective pollinations, with their thickness indicating their relative importance. To the left is the normal pin-thrum situation; the central column represents an evolving population; to the right is the endpoint with thrums eliminated.

On the basis of a mathematical model of the breeding system, it was possible (Crosby 1949) to illustrate curves of population change which agreed broadly with the range of population constitutions found in the homostyle area. With two important modifications, arising from a better knowledge of the populations, the same model is retained in this paper.

It is now known that the seed set is not the same for the three forms, and on the basis of results to be published later the ratio of seed set per plant for pins, thrums and homostyles is here taken to be 5:4:6, respectively. It is not surprising that the ready self-fertilization of homostyles should lead to a greater seed set, but this has relatively little effect on the course of the evolutionary change and on its endpoint; it does, however,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

roughly double the rate of evolution as compared with the rate derived from the 1949 model. The modified breeding matrix is set out in an appendix (table 2), together with a general indication of the computing techniques used in this paper.

In the 1949 model, no account was taken of the length of life of the primrose, the calculations being made at intervals of a generation, and with the assumption that all plants in a population were of the same age, and that all reproduced simultaneously in the year of their death. Results from a detailed study of two primrose populations have suggested that the average life of a primrose plant may be between 15 and 25 years. A flowering life of 16 years for all plants has been taken here, plus an initial non-flowering year of development; the population size remains constant. Calculation must now be carried out on a yearly basis; with mechanical computation this would mean a very great increase in the time required for the calculation of an evolutionary sequence (i.e. from the introduction of a homostyle to the elimination of thrums). The high-speed electronic computer is

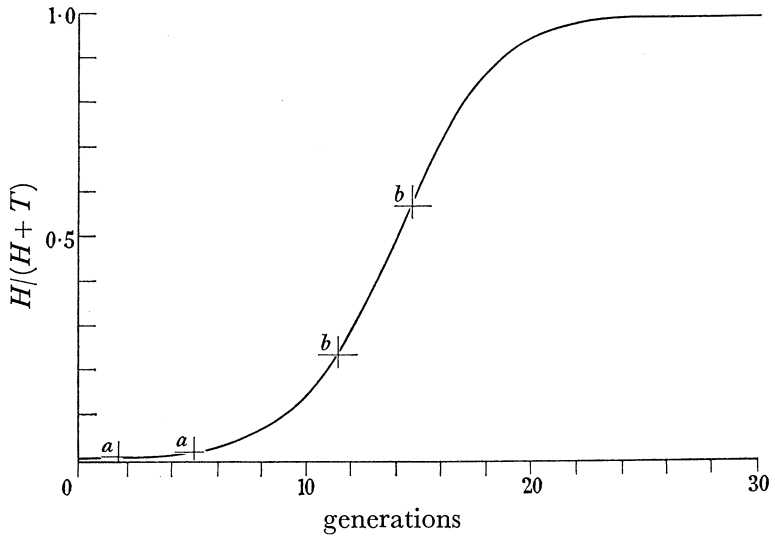


FIGURE 2. The calculated curve of the evolutionary sequence; homostyle frequency (as homostyles divided by homostyles plus thrums) is plotted against time in generations (1 generation = $8\frac{1}{2}$ years). *aa* and *bb* are sections covered in equal times, but with very different changes in homostyle frequency.

especially well adapted to iterative calculations of this kind; a complete evolutionary sequence can be calculated in about a minute, as against several days by the use of an electrical desk calculating machine.

One assumption made in 1949 was that heterozygous homostyles ($s's$), on selfing, segregate at least one quarter pins (ss), any deviation being due to relative inviability of the homozygotes ($s's'$). However, it now appears that this is not always true, and that such selfing frequently produces a shortage of pins (Crosby 1959). But there is much variability in this, and its relation to the progress of evolution to homostyly is uncertain; there does not at present seem to be any basis on which it can confidently be introduced into the mathematical model, and no further account can be taken of it in this paper. It will have to be considered eventually, when more experimental evidence has been obtained.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

A convenient way of expressing graphically the progress of evolution to homostyly is to plot against time the fraction of homostyles among high anthered plants, i.e. the fraction homostyles upon homostyles plus thrums, $H/(H+T)$. This has the advantage that it has precise endpoints at 0 and 1, whereas homostyle percentage, which might seem to be a more natural measure, has a fluctuating upper endpoint. For simplicity, the fraction $H/(H+T)$ will be referred to as homostyle frequency. In the maps its value will be represented by the degree of blackness of the symbols.

Figure 2 shows the progress of $H/(H+T)$ with time, beginning with a population containing one heterozygous homostyle in a thousand plants. Change is slow at first, but rapidly increases. With this size of population the endpoint is reached with the elimination of thrums in about 250 years.

Homostyle distribution in natural populations

The pattern of populations to be expected in nature is that of a fairly regular gradient of homostyly falling from a centre of high frequency, corresponding to an outward spread of homostyly from its point of origin. The gradient would not of course be quite regular, since variations in the density of primrose occurrence would naturally cause variations in the rate of outward flow of the homostyle gene complex.

A first glance at the map of homostyle distribution seems to contradict this expectation, for there are striking discontinuities in homostyle frequency; in some cases a population with no homostyles may be found quite close to a population with no thrums. There are many such local discontinuities, but close examination shows that they tend to occur most frequently across fields, whereas along lanes and hedgerows there is either uniformity or a recognizable gradient. This strongly suggests that the present pattern of homostyle distribution has arisen since the establishment of the present pattern of land usage, because it is clearly related to the fact that while primroses occur along hedgebanks, roadsides and lanesides, they do not normally occur in fields except close to boundaries. That is, while hedges, lanes and banks may form ready routes for homostyle spread, cultivated fields and pastures are barriers.

A full account of the detailed distribution of homostyly, based on over a thousand populations, will be published separately, but one typical illustrative example will be given here. Figure 3 shows on a large scale a small section of the distribution map for the populations in Somerset. There are obvious irregularities in homostyle distribution, but it will be seen that these are more pronounced in the south-west quarter of the map, where primroses are less frequent and tend to be local. In the centre, north-east, and east, where they are very common and well dispersed, there are only small differences in homostyle frequency over large areas; in particular, there should be noted two lanes in which primroses are very abundant, and a strip of woodland running eastwards, where they are also very common.

We may imagine a simple situation, with a population containing homostyles, and from which there radiate a number of routes *along* which homostyly may spread because of the presence of primroses, but *between* which such spread is impossible because of their absence. Difference in homostyle frequency on opposite sides of a field might then be explained by differences in the rate of spread of homostyly along the radii, resulting

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

primarily from differences in primrose density. But in any particular instance the effect of this would be difficult to estimate, because the present density of primroses is by no means a reliable guide to past density. My general feeling from a study of a number of similar situations is that differences in the ease with which homostyly can spread can only partly explain the irregularities which occur in homostyle distribution. Another possible cause might be random fluctuations in the frequencies of the various genotypes, and this possibility will now be examined.

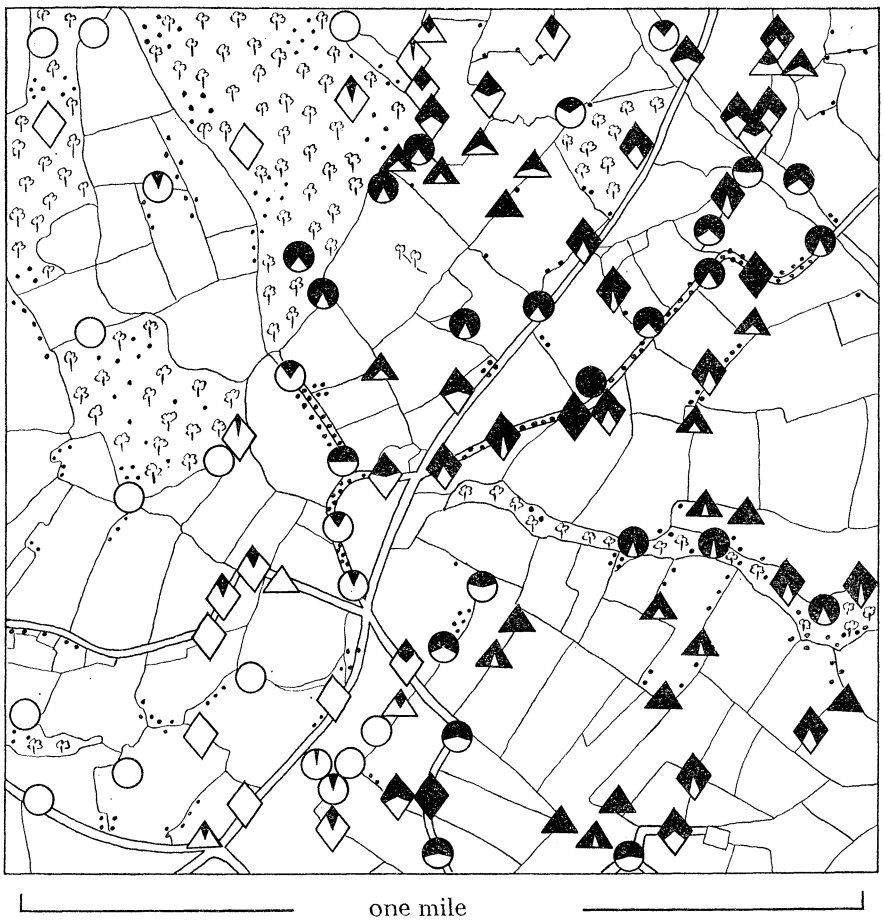


FIGURE 3. A small part of the general distribution map of homostyly in Somerset. Single lines are usually hedges; double lines are roads or lanes; trees indicate woods. The shape of the population symbols indicates the size of the sample scored for that population; for sizes less than 90, this will give a fair indication of the size of the population; the sample sizes are △, 11–40; ◇, 41–90; ○, over 90. Prominent black dots indicate where at least a few primroses occur but have not been scored. The degree of blackness of the symbols indicates the value of $H/(H+T)$ (see figure 5); white symbols have no homostyles, black symbols no thrums.

RANDOM FLUCTUATIONS AND THE EVOLUTIONARY SEQUENCE

Referring to figure 2, it can be seen that a small fortuitous increase in the number of homostyles in a population near the beginning of the curve would have the effect of causing that population to gain several years of evolutionary progress. Subsequent fluctuations might reduce or increase this gain, but even if it were simply maintained it is clear that by the time the steepest part of the curve is reached, that gain in years would be

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

translated into a substantial difference in homostyle frequency; this can be seen by comparing the points *aa* and *bb*. This strongly suggests that two identical populations starting with a low homostyle frequency might come to differ considerably in their constitution after, say, 150 years, simply as a result of annual fluctuations superimposed on a rapidly increasing rate of evolution.

The direct general mathematical estimation of the probability distribution of such differences is a formidable task, especially since four genotypes are involved, and it may be impossible. But it may be approached indirectly by superimposing normal variation on the evolutionary sequence in a series of calculations. Each such calculation would then represent the history of a single isolated population, and repetition would allow the comparison of a number of imaginary populations in respect of differences in their evolutionary history which have arisen solely as a result of chance fluctuations in gene frequency. Electronic computation allows the performance of such a task, which otherwise would require a time far beyond the limits of practicability.

The process which has been used will be outlined here; a more detailed account is given in the appendix. The variability is superimposed with every annual calculation. For each year, from the values p_0, q_0, r_0 and s_0 for the frequencies of the four genotypes in the population, there are calculated (from equations derived from the breeding matrix shown in table 2) the expected frequencies p_1, q_1, r_1 and s_1 for those genotypes among the offspring produced by the population in that year. These frequencies are considered as though they were the frequencies in a large population, from which the computer then takes a random sample of the required size, using for this purpose calculated sequences of pseudorandom numbers which may in practice safely be considered as random. That is, the new plants to be incorporated in the population are not necessarily in the calculated frequencies, but in frequencies which might have occurred naturally through the roughly random sampling processes of reproduction. The iterative calculation of the evolutionary sequence then proceeds from the actual and not the expected new population constitution. When one evolutionary sequence ends with the final elimination of thrums (or fails to start properly with the fortuitous elimination of homostyles), another sequence is started automatically.

For each of six different population sizes, starting with the introduction of a single heterozygous homostyle, forty such evolutionary sequences have been calculated. In order to save computing time, for each sequence the population constitution was printed only at intervals—once every two generations. The total computing time occupied by the whole operation was about 8 hours; other methods of calculation would have required several years.

As an example, the first twenty-three runs for population size 256 (adult plants) are shown in figure 4. Thirteen of these failed because of early elimination of homostyles, and only ten were complete runs with eventual elimination of thrums. The dotted line shows the theoretical curve without superimposed fluctuations.

Of the failures, in ten the initial homostyle failed to leave any homostyle descendants, and the population relapsed to full heterostyly with its death at 17 years. In two of the other three, one new homostyle was produced by the time the original one died, while in the third there were two new homostyles by this time; nevertheless, all of these failed to leave any homostyle descendants, and by year 34 there were no homostyles in the populations.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

In the populations which made progress, recording at 17-year intervals has obscured yearly fluctuations, but general trends are easier to follow. It is clear that after only a few generations there were great differences between these populations, as had been expected. At generation 10 (85 years) the range in value of homostyle frequency in these ten populations only was from 0·148 to 0·843, or from 8·2 to 58·6 % of homostyles in the population. The sequence was completed in times ranging from sixteen to twenty-eight generations; the last population to finish (well behind the rest) was actually lying fifth after six generations. The two extreme populations at generation 10 had been equal sixth at generation 2, and finished six generations apart.

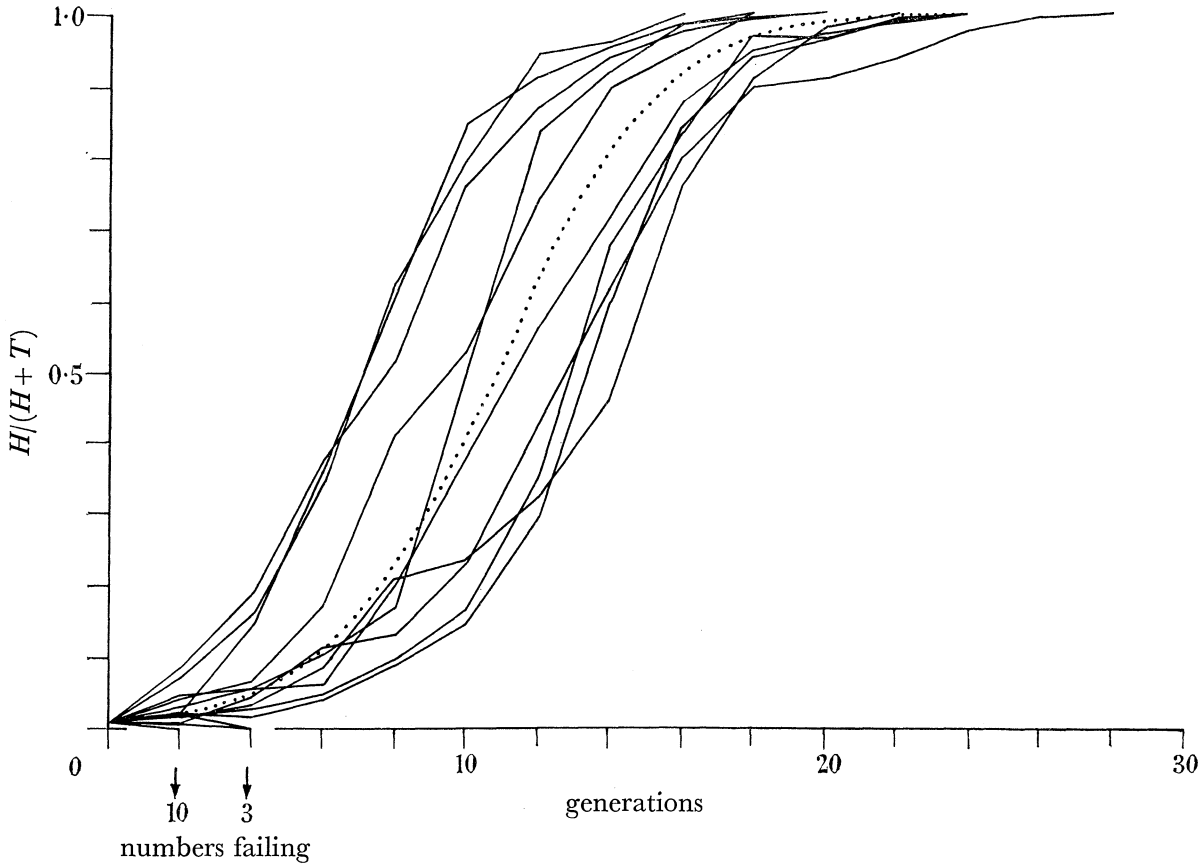
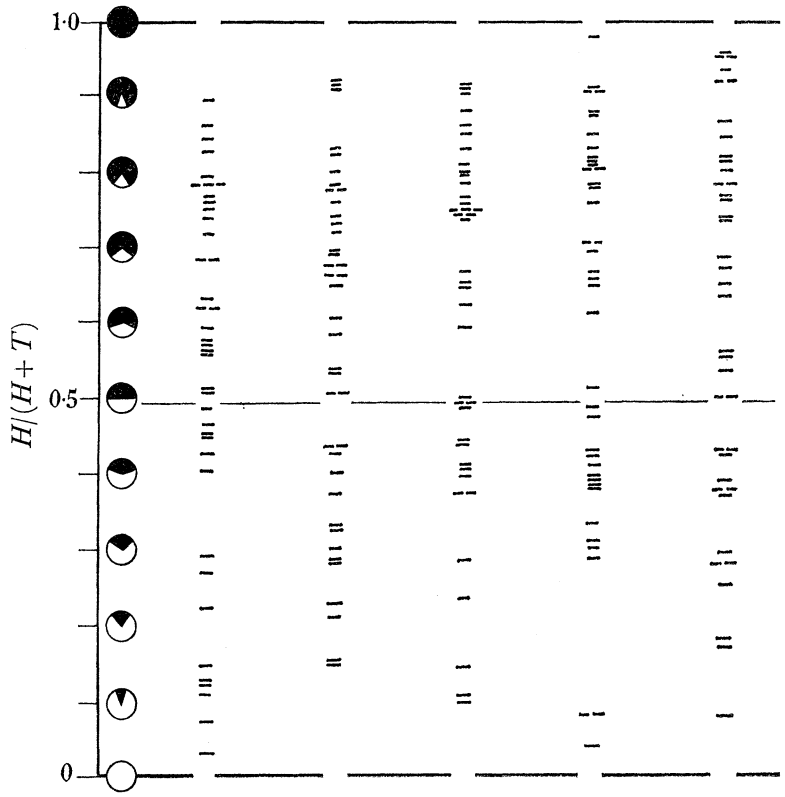


FIGURE 4. The effect of random fluctuations on the evolutionary sequence. The curves give, at 17-year intervals, the homostyle frequencies of the first ten populations of 256 flowering plants, initially identical and each including a single homostyle, which made permanent evolutionary progress. Thirteen failures are indicated. The dotted line gives the theoretical curve without fluctuations.

It must be emphasized that these are not selected curves, but are the first ten completed sequences to come from the computer at this population size. Other population sizes show similar results, except that the curves tend to be smoother in the larger population (1024 plants) and more irregular in the smaller ones, as might be expected. Some of the more striking results include a population of 128 plants which at successive two-generation periods from the start had 2, 1, 1, 1 and 0 homostyles. At population size 32, three out of forty sequences were completed in six generations, while one took thirty; and one of the

failures kept going for eight generations before losing its homostyles. In the smallest population (sixteen plants), thrums were sometimes eliminated in four generations, but the most striking population was one of this size in which there were five homostyles by the second generation, but none at all by the tenth—a notable example of the Sewall Wright effect. The population sizes which have been chosen for this work are reasonably representative of natural populations over most of the area which has been studied.



population size	1024	256	128	64	32
years	119	91	77	62	48
failures	40	29	31	27	29

FIGURE 5. The effect of random fluctuations on the evolutionary sequence. Each column gives for one population size the homostyle frequencies of forty evolving populations at a time when, without fluctuations, they would all have been expected to have a value for $H/(H+T)$ of 0·49 (indicated by a thin horizontal line). Population symbols as used in the maps show homostyle frequency by their degree of blackness, and indicate the progress of evolution. For each population size there is given at the foot of the figure the time allowed for each of its forty runs, and the number of occasions on which homostyly was lost before the fortieth successful population appeared.

A general survey of the results is shown for all but the smallest population in figure 5. Each column of points represents forty evolving populations of a particular size, and each point shows the state of evolution reached (the value of $H/(H+T)$) at a time which is constant for each population size, but differs between the different population sizes in the following way. For each size, the time chosen is that at which a value of 0·49 for $H/(H+T)$ would have been reached had there been no fluctuations. This time differs for the different sizes because, since each population starts with a single homostyle, the smallest populations

start with the highest homostyle frequency, and thus reach the value 0·49 earlier. It might be thought that it would be more reasonable to compare them after the same time interval, but this is not so. Gene flow probably takes place more readily into large populations than small ones; so although a smaller population will start with a higher frequency of homostyles, it will usually start later, and these two things may tend to cancel out.

0·49 was chosen as being about half-way through the sequence, and the value expected for the largest population at the seventh printing. The results from the original set of forty sequences for this population size could then be used. For the other four population sizes, new sequences were run off, but only for the number of years required and only the one result was printed for each sequence.

Population symbols as used in figure 3 indicate the stages of the evolutionary sequence, and the horizontal line through figure 5 is drawn at 0·49. At the foot of each column is indicated the number of populations which failed to make permanent progress and lost their homostyles. In nature, of course, some of these might have started again by the influx of further homostyle pollen or seed, and would appear as points near the bottom of the columns; but how soon this would happen would depend upon several factors including the degree of isolation of the population; the results as shown are sufficiently striking, and do not need this refinement.

Three things stand out immediately from figure 5. The variation about the 'expected' value is not even approximately normal, but this is partly because the centre of the column is the region of most rapid evolution, and therefore most rapidly passed through. The variation is roughly the same for all population sizes; this is not so surprising as it seems to be at first—it happens because the larger populations have already varied considerably by the time they have reached the uniform starting points of the smaller ones; for example, the 64-plant populations all start at a value for $H/(H+T)$ of 0·033, but this value is not 'expected' for the largest population until after about 50 years, by which time the forty runs there already show a range of variation from 0·002 to 0·146.

Finally, it is clear even from the largest population, with its range after 119 years from 0·026 to 0·891 (from 1·3 to 62·0 % of homostyles), that random fluctuation within populations could be quite sufficient by itself to account for the large irregularities in homostyle distribution which are found in nature, provided the populations are sufficiently isolated.

Random fluctuations and gene flow

Random fluctuations can enter the problem in a different way. It has already been pointed out that discrepancies between adjacent radii could partly be accounted for by different rates of movement of the homostyly gene-complex along the radii, due to differences in primrose density. But this is also a process in which chance can play a large part, and it is worth while considering the effect this would have on a group of populations, in conjunction with the intrapopulation effect which has just been considered.

It is first necessary to establish a basis for gene flow. There are only two ways in which this can happen naturally—by pollen or seed. Primrose seeds are rather large, and are produced in capsules which are tucked under the leaves; the possibility of purely mechanical dispersal is negligible. They are, however, sweetly sticky with an aril, and this is very

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

attractive to ants which remove them from the capsules, and which are probably the main dispersal agents.

It is very difficult to assess how far seeds may be taken in this way, but one would not expect it to be very far. This is supported by observations on seedlings, which are common enough near primrose plants, but whose frequency falls off rapidly (within decimetres) away from plants. The tendency for primrose populations to have sharp edges, and the relative infrequency of isolated plants between populations, also suggest that while gene flow by seed dispersal undoubtedly occurs, it is much less important than gene flow by pollen.

The pollinating agents of the primrose have been a quite unnecessary mystery for a long time, and many people have failed to observe consistent insect visitors. In Somerset at least, the primrose is regularly pollinated by bee-like diptera, such as *Bombylius*; these move very rapidly, and one can pollinate many flowers in very few minutes. The quickness with which the job is done accounts for the infrequency with which they have been observed. The brimstone butterfly is an occasional and more leisurely pollinator.

During the early days of research into homostyly, the occurrence within populations of considerable heterogeneity in phenotype frequencies had led me to believe that the breeding range of an individual plant was very small—which would account for such heterogeneity (Crosby 1948). It was this belief which led me to approach the present problem with the assumption that the transport of homostyle pollen from one population to another was a relatively uncommon event, and to plan the randomization technique accordingly. The method first used was thus to assign to each pair of adjacent populations a constant which represented the chance (per plant of the appropriate genotype in the donating population) that a plant would appear in one population which had been fathered by a plant in the other; this allowed for the flow of thrum and homostyle in either direction. The constant was determined by reference to the sizes of the two populations and their distance apart. The constant multiplied by the frequency of the donating genotype was then compared with a pseudorandom number to determine whether a new genotype should be added to the recipient population that year.

This method produced a very satisfactory simulation of the natural distribution pattern, but is almost certainly unsound biologically; it was therefore abandoned and the results rejected.

The objection arises from the observation that even a small primrose plant may produce more than fifty seeds per annum, and that owing to rotting of developing capsules and damage by animals such as mice, these ripe seeds represent a much greater number of fertilized ovules. Numbers will vary greatly, but a small plant may roughly be supposed to have 200 ovules fertilized per annum. If a primrose plant lives for 16 years, and if the population size remains constant, this means that one new plant corresponds to over 3000 functioning pollen grains, and to many more which may reach stigmas but fail to reach an ovule.

It must be concluded that pollen is carried fairly freely from one population to another. To consider gene flow as though it resulted from the occasional transport of a pollen grain from one population to another (which is really what the method outlined above amounts to) is therefore quite inaccurate.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

A better system is to consider that the male plants for any population consist of its own plants plus fractions of the immediately neighbouring populations, and this has been adopted here. These fractions are constants for any desired general rate of gene flow, and depend for each pair of populations on the distance apart and on the size of the recipient population. Basically, two types of gene flow have been considered—fast and slow. Fast flow has been imagined as taking place in an area where the primrose is relatively abundant, and the ease of gene flow falls off slowly with distance; slow flow is imagined as relating to an area where the primrose is less abundant and populations are more isolated; it falls off more rapidly with distance. For medium distances fast flow was four times as fast as slow flow, and the actual rates were chosen so that the outlying populations in the imaginary distribution maps about to be described were reached by homostyly in a reasonable time. The relationship between fast and slow flow is indicated in table 1, in terms of the fraction constant.

TABLE 1. THE RELATIONSHIP BETWEEN SLOW AND FAST RATES OF GENE FLOW

Each figure in the table gives that fraction of a population which is considered as donating pollen to a neighbouring population of 256 plants. The distances are the shortest and longest distances apart of neighbouring populations in the maps of figures 7 to 9, and have a ratio of 1:8.

	shortest distance	longest distance
slow flow	0·30	0·014
fast flow	0·50	0·21

The assessment of the effect of the size of the recipient population was largely guesswork. It seems reasonable to suppose that insects would find and visit a large population more readily than a small one. As a provisional working arrangement it was supposed that the chance of a population receiving pollen from its neighbours varied with the square root of its numerical size; with very large populations it is doubtful whether population size would make any difference.

The supposition that gene flow takes place only between immediately neighbouring populations cannot be completely true, but it was felt to be a reasonable simplification in a first approach to this problem. As a rule, insects would probably not fly over a population without visiting at least a few flowers, with consequent modification of their pollen load; the proportion of the pollen carried by an immigrant insect which is from non-adjacent populations may probably therefore be relatively small, though by no means negligible. To take such pollen into account will require a very much increased complexity of the calculations.

When, for any one population, the expected contribution in terms of pollen parents from all immediately neighbouring populations has been decided, random fluctuation is superimposed by the method described in the appendix. It is considered that this variation arises largely from differences in insect activity from year to year, and not merely from sampling errors in deriving the fractions of the donating populations (which would give a much smaller variability).

The first computations involving fluctuations in gene flow as well as in intrapopulation evolution dealt with the progress of homostyly along ten identical and isolated lanes, five with fast gene flow and five with slow (figure 6). No particular scale is laid down for these

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

lanes, but they may be imagined to be about a mile long. From left to right, the population sizes are 256, 128, 128, 256, 128, 128 and 64.

Each lane starts with one heterozygous homostyle, which is in the population at the left-hand end. If any lane fortuitously loses its homostyles, the run is ignored and the lane is restarted at year 0. This has the effect of reducing the heterogeneity of the results, but the only alternatives were to let the lane remain permanently free of homostyles, to restart with one homostyle at the year of failure, or to restart at some later year. The application

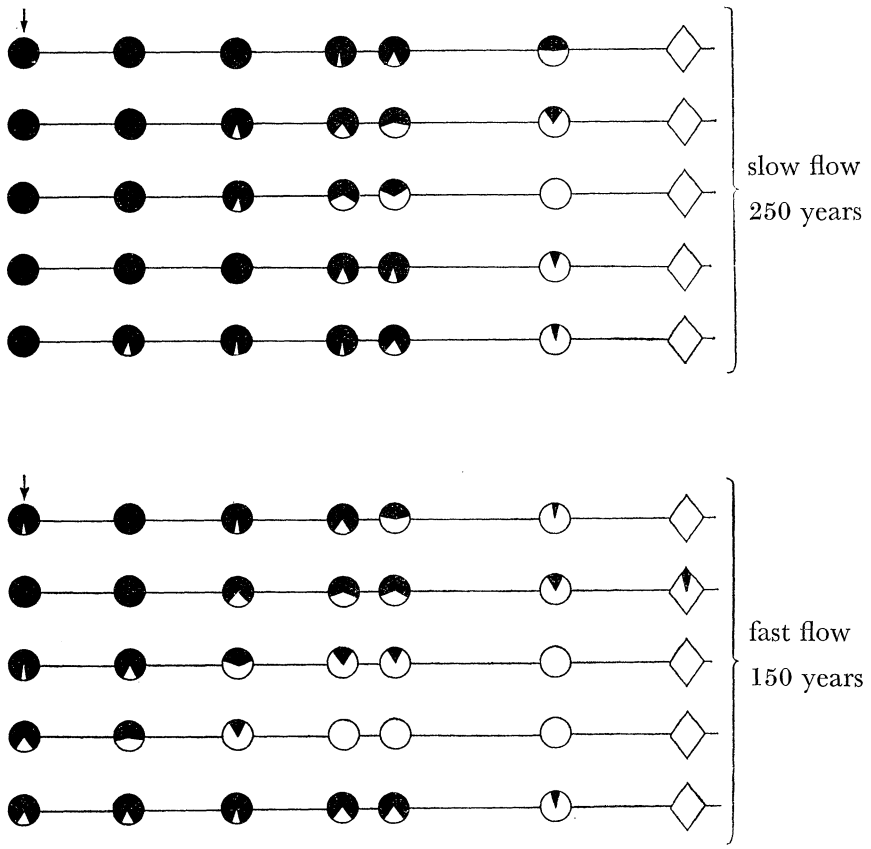


FIGURE 6. Random fluctuations and gene flow. Ten identical lanes with primrose populations are represented here. At year 0 the initial population of each lane (indicated by arrows) contains a single homostyle, the other populations none. The top group shows the distribution of homostyly in five lanes with slow gene flow after 250 years; the bottom group shows five lanes with fast flow after 150 years. Population symbols as in figures 3 and 5.

of any of these alternatives would be quite arbitrary, and it seems better to discount the failed runs, but to remember this when considering the results. Of the five lanes with fast flow, two failed once and one twice; of the five lanes with slow flow, only one failed, but did so three times.

The runs were of 250 years, population constitutions being printed every 50 years. The results are illustrated in figure 6 which shows the slow flow position after 250 years and the fast flow position after 150 years. Between the five lanes of each group, the differences are due solely to chance. A comparison of corresponding populations shows that chance alone, operating on evolution rate and on gene flow, can adequately account for the discontinuities in homostyle distribution which are found in nature.

The differences between the two sets were not as great as might have been expected, except for obvious direct consequences of difference in flow rate. The gradient of homostyle frequency is necessarily steeper in the case of slow flow, but there is little more than a suggestion of greater discontinuity along the lanes near their ends than in the fast series. Differences between the lanes are more marked in the case of fast flow, and this is contrary to the expectation that chance would have a greater effect on the lower frequency of homostyle migration; five runs may not be enough to allow a good comparison between the flow rates, and the results must be considered in conjunction with the next series.

Although the distance progressed by homostyly along the lanes must to some extent depend upon the early rate of accumulation of homostyles in the first population, that this is by no means the whole story can be seen by reference to the fast series. For the fifth population, the bottom lane has the highest homostyle frequency after 150 years, although at 50 years the first population in this lane had the lowest frequency. However, in spite of this low frequency in the initial population (six homostyles in 256 plants), homostyly had by this fiftieth year reached the fourth population in the lane, whereas in no other lane had it progressed beyond the second.

The investigation of gene flow was next extended to a more complex situation. An imaginary map was drawn; no precise scale was intended, but it may be imagined to cover an area of between one and two square miles. Thirty-seven populations are considered at various distances apart, and of sizes ranging from 32 to 1024 plants; they are situated on lanesides, hedgebanks, by streams, and in woods. Precisely the same treatment was followed as in the single lane comparisons, with the proviso that while gene flow could take place along lanes, hedges, streams and through woods, it could not do so across fields. As before, gene flow takes place only between immediate neighbours. The map has four branching lines of gene flow radiating out from one central population (indicated by an arrow). There is no cross-flow between the radii, except at the ends of one adjacent pair; in nature, some lines of cross-flow would be expected, but it was felt that it would be more instructive in the first place to omit them. Each run is started by placing a heterozygous homostyle in the central population; if homostyly is lost from the map, the initial conditions (except the pseudorandom number sequence) are automatically reset and the process restarted from year 0.

For each flow rate three runs of 250 years were made, with printings at 50-year intervals. These six runs and the ten single-lane runs were carried out in one continuous operation (lasting about 15 hours), during which the pseudorandom number sequence was maintained without interruption.

Considerations of space forbid the presentation of all the results, but as illustrations two fast-flow positions after 200 years are shown in figures 7 and 8, and figure 9 shows a slow-flow position after 250 years. In general, the pictures which these maps present bear a satisfactory resemblance to that presented by figure 3, with similar uniformities, gradients, and discontinuities. They confirm and amplify the earlier conclusions on the part played by random fluctuations in the establishment of the present pattern of homostyle distribution in Somerset.

The slow-flow runs all showed more marked discontinuities towards the ends of the radii, but there was no consistent difference between the flow rates in respect of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

interradial discontinuities, some of which were very considerable, as can be seen from the figures.

In general, the tendency is for the slow-flow maps to resemble the south-west quarter of figure 3, where primroses are less common and more localized, and gene flow is probably less free. The fast-flow maps tend to resemble more closely those parts of figure 3 where the primrose is abundant and continuously distributed. Particularly striking is figure 8; there is a connected line of populations with almost uniform homostyle frequency, running from

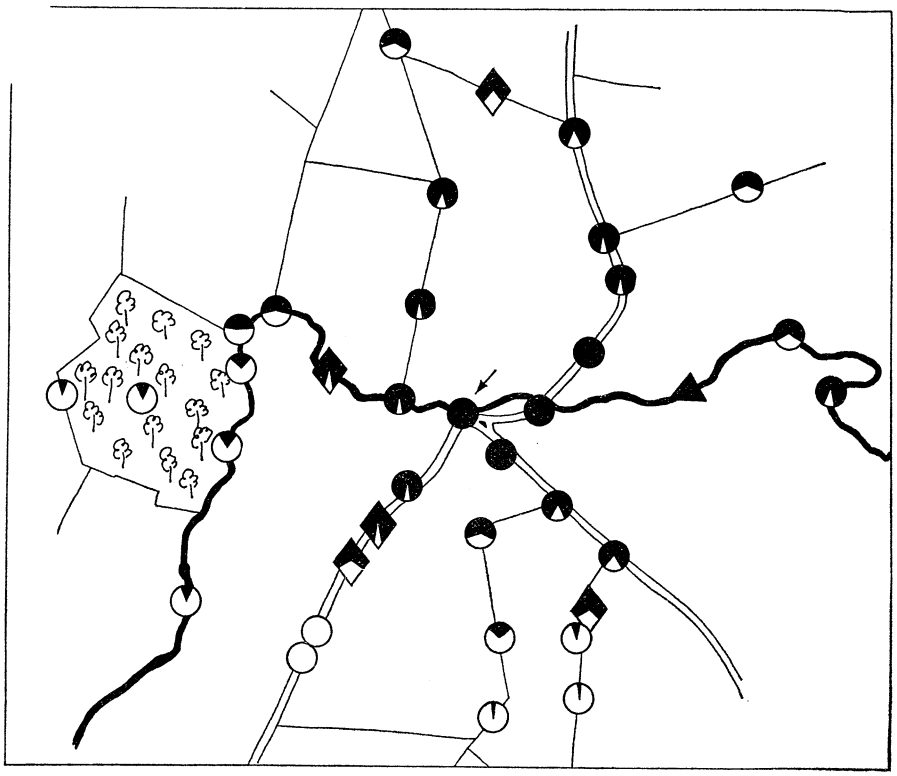


FIGURE 7. An imaginary map with primrose populations. Population and map symbols as in figures 3 and 5; the thick line is a stream. No precise scale is intended, but the area may be imagined to be 1 to 2 square miles. At year 0, there was only one homostyle in the area, and that was in the central population (indicated by an arrow). This figure shows the distribution of homostyly after 200 years, with fast gene flow.

top centre to bottom centre of the figure; this should be compared with the lane running from the centre of figure 3 towards the north-east corner. Figure 8 also shows three cases of a reversed gradient of homostyle frequency.

The equal occurrence of large interradial discontinuities in the fast- and slow-flow series arises at least partly from the fact that the basis for distinguishing slow and fast flow which has been used here is somewhat artificial. In nature, areas of fast gene flow, being areas with abundant primroses, would have interradial connexions which have been almost entirely disregarded in these imaginary maps; where these occur, they will allow gene flow between adjacent radii with a consequent levelling of the homostyle frequencies.

Slow and fast flow would thus better be compared on different basic maps, with provision for interradial gene flow in the case of fast flow. It is intended to deal with this by

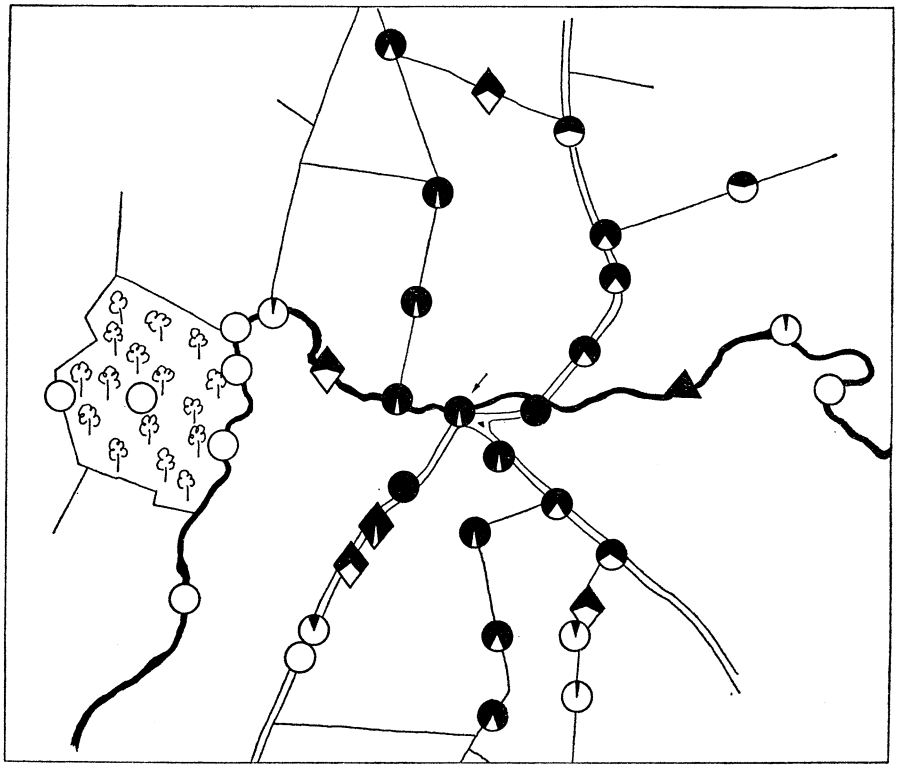


FIGURE 8. See caption to figure 7. This figure shows the distribution of homostyly after a second run of 200 years, with fast gene flow.

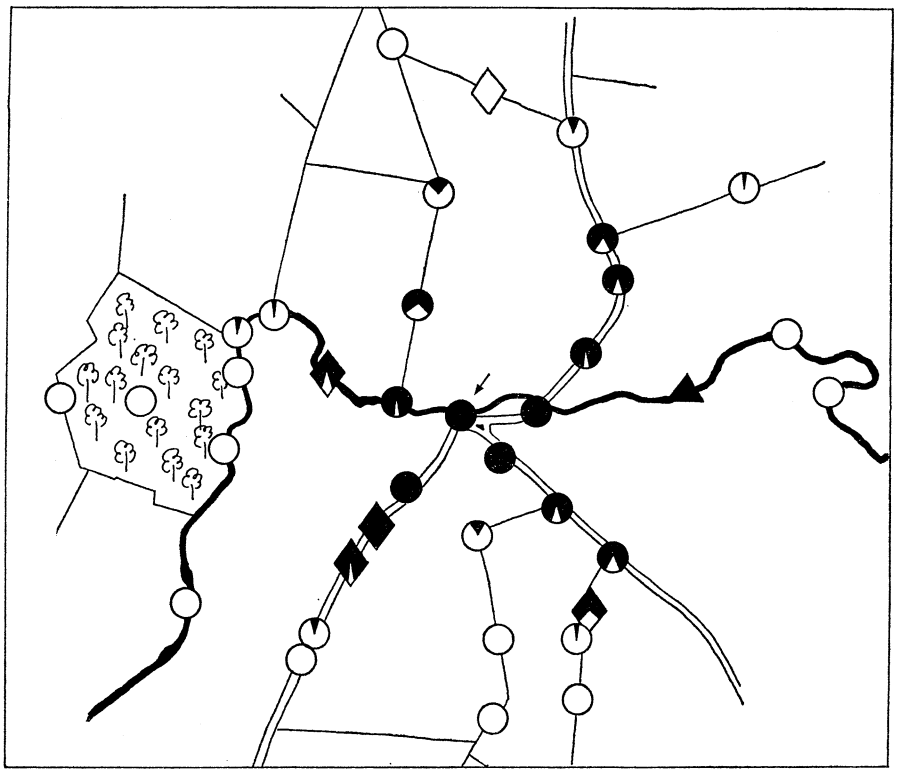


FIGURE 9. See caption to figure 7. This figure shows the distribution of homostyly after 250 years with slow gene flow.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

considering a larger imaginary map, with over a hundred populations, and with the two flow rates on different parts of the map and determined solely by primrose frequency and distribution. It is also proposed to extend the range of gene flow by allowing pollen transport to the next-but-one as well as the next population.

A comparison of figures 7 and 8 illustrates the differences which could arise by chance between two identical sets of populations with identical gene-flow conditions. Three of the radii are quite different in the two cases (but not always in the same direction); this was much more pronounced 50 years later, especially in the wood region on the west side; in one case (from figure 8) homostyly had only just reached the wood by year 250, and was absent from the four distal populations; in the other (from figure 7), no populations in that region had a value for $H/(H+T)$ of less than 0·76.

Similar differences appear between the three slow flow sets. One of these at 250 years even shows more homostyly in the wood than the fast flow 250-year map following from figure 8.

Discussion

Although the primary purpose of this work was to explore the potentialities of electronic computation techniques, it has already made a useful contribution to the study of the homostyly problem. Where facts were lacking the programming had to be based on assumption; gaps in knowledge were thus emphasized and their relative importance indicated. This was particularly so in the field of gene flow, where it has been necessary to revise completely my earlier ideas about pollen transport in the primrose.

Examples of considerable heterogeneity within populations containing homostyles had led to the conclusion that the breeding range of the individual plant was very short, since this could easily account for the heterogeneity observed. But in the artificial maps the simulation of the natural distribution pattern required the assumption of a very free flow of pollen between primrose populations, and the idea of a short breeding range was abandoned. The detailed distribution of homostyly will be discussed more specifically in a later paper, but the whole distribution picture, which occupies a very much larger area than the map of figure 3, would be very difficult to interpret without the assumption of free pollen flow. It would be necessary to assume occasional new occurrences of homostyly (by crossover or mutation), or that the rate of evolution is very much slower than that suggested in this paper, or that there is a considerable slowing down of the rate of evolution in the later stages; there are theoretical objections to the first of these alternatives, and the facts seem to be against the others.

The assumption of free pollen flow does not seem to be open to any serious objections, nor is there anything improbable in the suggestion that pollen may occasionally be carried several miles; this cannot of course be tested over the small imaginary areas of this paper, but its assumption would remove the main difficulties in the interpretation of the distribution of homostyly in Somerset. It is true that in his studies of crop contamination in certain insect-pollinated plants, Bateman (1947*a*) found that the amount of pollen transported fell off rapidly in quite a short distance, although a very small proportion of pollen might travel over much longer distances. Although not stated explicitly, it can be inferred that the pollinating agents were mostly bees, and in a later paper (1947*b*) discussing these results he deals almost exclusively with bees. But bees are probably of little importance as

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

primrose pollinators in Somerset, and it is species of *Bombylius* which have to be considered. These, with their rapid darting flight, and the ease with which they are alarmed, probably have a different pattern of behaviour from bees in their flower-visiting flights, and this may produce quite a different pattern of pollen distribution; it is possible that differences in the relative importance of different groups of pollinators may explain the apparent differences in the pattern of homostyle distribution found in Somerset and Buckinghamshire.

It has not been possible to find very much relevant published information about *Bombylius*, and it is clear that before the pattern of homostyly distribution can be fully understood a detailed study must be made not only of the flight behaviour of *Bombylius*, but also of its general natural history. As knowledge accumulates, the provisional assumptions made in this paper can be replaced or made more accurate, and the computer models will be used to explore the implications of the new knowledge and to suggest the most profitable directions for further inquiry.

From a consideration of the entire homostyle area in Somerset, it is tentatively suggested that the rate of spread of homostyly is of the order of 5 to 10 miles in about 100 years. Although this is much faster than the calculated spread in the imaginary areas, it must be remembered that these were not allowed occasional long-distance transport of pollen, and that pollen of one population was not allowed to pass beyond the next neighbouring population. The removal of these restrictions would have greatly increased the rate of spread of homostyly.

The extent of pollen movement which has been allowed in this paper can be indicated by reference to the fast-flow calculations, which more nearly represent the natural conditions. It has been reckoned that for two populations at about one-sixth of a mile apart (about one-quarter km), approximately one-fifth of the plants in one population may be regarded as contributing pollen to the other; for populations very close together this fraction may be well over one-half.

Intrapopulation heterogeneity still requires explanation, although later counts showed that it occurs much less frequently than had been supposed and that the earlier results were misleading. It seems possible that examples of it may have resulted from sudden increases in population size, which are known to occur in *Primula vulgaris*. It is, however, proposed to test a model population, constructed in the computer in terms of individual plants in specific positions in the population, to see whether heterogeneity could arise without sudden size increase or other occasional fortuitous event.

It is hoped that this paper has demonstrated that the electronic computer is a potentially valuable tool for the elucidation of complex evolutionary situations involving an element of randomness. In particular, it provides a method of investigating problems of gene flow which could not otherwise be tackled; the examples given here are relatively simple, and do not by any means indicate the limits of the technique.

Successful use of electronic computation has been made by other biologists working in a number of different fields. Among others, mention may be made of the ecological studies of Williams at Southampton (referred to by Williams & Lambert 1959) and the genetical studies of Fraser (1957), and several American biologists are using computer techniques.

It is hoped that more biologists may be encouraged to turn to the possibility of using electronic computation in the elucidation of their problems. It is particularly urged that

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

they should learn for themselves the art of programming, so that they can handle their problems themselves without the need to call in mathematical help. The learning required is not particularly difficult, nor does it require any advanced knowledge of mathematics; the basic processes of the computer are arithmetically simple, and it does not need the refined techniques of higher mathematics. Only someone with an understanding of the manipulative possibilities of the computer and of the biology of the system being studied can use the one to the best advantage in the elucidation of the other. It is easier for a biologist to learn what a computer can do than for a mathematician trained to abstract thinking to understand the reality of biological situations.

The complexity of such situations has in the past necessitated much simplification before they could be treated mathematically, and such simplifications made without a proper understanding of biological principles are very liable to lead to unreality. Mathematical abstractions of biological situations, although admirable in themselves, may sometimes lack real validity. This has been to some extent inevitable, because time has often been a limiting factor in such work. But time becomes a factor of less significance when electronic computation is available; moreover, the machine can deal easily with very complicated situations, and the dangers of simplification are thereby lessened. In the hands of a biologist the computer can allow a much more realistic approach to the complex problems of biology than has hitherto been possible.

I would like to express my thanks to the Director (Dr E. S. Page) and the Staff of the University of Durham Computing Laboratory for their encouragement and friendly assistance, particularly when I was learning to use their machine. I gratefully acknowledge generous financial grants towards the field work on the homostyle populations which have been made at various times over a number of years by the Royal Society, the Nature Conservancy, and the Council of the Durham Colleges.

Appendix. An outline of the computing methods used

This work was carried out on a Ferranti 'Pegasus' computer, which has a main store capacity of 512 accessible eight-word blocks, and a computing store capacity of six eight-word blocks plus seven single-word accumulators. The computing store is just big enough to handle the programs economically with regard to minimizing block transfers between the stores. The computer works on a binary scale.

The evolutionary calculations

The breeding matrix which is the basis of the calculations is shown in table 2. This corresponds to table 1 of Crosby (1949), modified by allowances for different seed productivities and by the consideration that the genotype frequencies of male parents are not necessarily the same as those of female parents, except in an isolated population.

The calculations are made on a yearly basis. For isolated populations, it is assumed that each plant lives 17 years, not flowering in the first year, and that the number of plants is the same for all age groups (that is, the same number of plants die each year and are replaced by an equal number of new plants). Sixteen years was chosen as the length of flowering

life, being the most appropriate round number in the binary scale. The average generation time is then about $8\frac{1}{2}$ years.

The assumption of no variation in length of life is of course not correct, but at present there is not enough information to allow even a rough guess at the variability in this respect. It is unlikely that this has had any important qualitative effect on the conclusions reached; any increase in precision which might result from taking such variability into account would probably not be worth the considerable complications of the calculations, but this is a point which will be investigated separately.

TABLE 2. THE BREEDING MATRIX OF HOMOSTYLE-CONTAINING POPULATIONS

p, q, r, s are the frequencies respectively of pins, thrums, heterozygous homostyles and homozygous homostyles in the population.

b, c, d are the frequencies respectively of thrums, heterozygous homostyles and homozygous homostyles among the pollen parents for the population; in an isolated population their values will be the same as q, r, s.

v is the viability of homozygous homostyles relative to the other three genotypes.

fertilization	frequency	relative seed set per plant	proportions of genotypes among the offspring			
			ss	Ss	$s's$	$s's'$
$ss \times ss$	$\frac{p}{10}$	5	$\frac{p}{2}$	—	—	—
$ss \times Ss$	$\frac{9pb}{10(b+c+d)}$	5	$\frac{9pb}{4(b+c+d)}$	$\frac{9pb}{4(b+c+d)}$	—	—
$ss \times s's$	$\frac{9pc}{10(b+c+d)}$	5	$\frac{9pc}{4(b+c+d)}$	—	$\frac{9pc}{4(b+c+d)}$	—
$ss \times s's'$	$\frac{9pd}{10(b+c+d)}$	5	—	—	$\frac{9pd}{2(b+c+d)}$	—
$Ss \times ss$	q	4	$2q$	$2q$	—	—
$s's$ self	r	6	$\frac{3}{2}r$	—	$3r$	$\frac{3}{2}r.v$
$s's'$ self	s	6	—	—	—	$6s.v$

The starting point for each evolutionary sequence is the equilibrium pin : thrum constitution, with one thrum replaced by a heterozygous homostyle. The population is then divided into seventeen portions with equal totals, and as nearly equal in constitution as the maintenance of whole numbers permits. These portions are stored in successive blocks ('year blocks') of the main store, four successive words of the block carrying the numbers of the four genotypes; the last block is considered the youngest, and contains the homostyle. A further block ('population block') contains the total numbers of flowering plants for the whole population.

When the new plants produced in the first year have been determined (as explained later), the plants in the first (oldest) year block are killed by subtracting them from the totals in the population block, and by replacing them in the year block by the new plants; the plants in the youngest year block are now added in to the population block—that is, there is a year's lag representing seedling development with no flowering. The first year block then becomes the youngest, and after the second calculation the process is repeated with the second year block, and so on. After 17 years all the original plants have been replaced, and the cycle returns to the first year block and is repeated as many times as required.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

For the calculations introducing gene flow it was necessary to use computer storage space economically, in order that many populations could be dealt with simultaneously. It was found that the introduction of an initial non-flowering year had made no significant difference to the results, except that it slowed the rate of evolution somewhat. By dropping this refinement, it was possible to store the remaining 16 year groups in eight blocks, one group occupying either the first or the last four words of a block; the 16-year cycle proceeds in half-block steps. It was then possible to store simultaneously thirty-seven populations in the computer. Together with the programme (152 blocks), these occupied almost the whole of the available storage space. [It would, however, be possible, by packing the year groups into single words, approximately to quadruple the number of populations which could be dealt with simultaneously; the computing time for a 250-year run would then be about 10 hours.]

The randomization process

Essentially, the method by which fluctuations are superimposed on the evolutionary sequence is that each year, when new frequencies p_1, q_1, r_1 and s_1 for the four genotypes have been calculated on the basis of table 2, the machine is then made to behave as though it were selecting from a large population with these frequencies a random sample of the size required (one-sixteenth of the flowering population size). This sample will be the year's new plants previously referred to.

The sampling process is carried out with the aid of sequences of numbers generated through calculation by the computer. Such numbers cannot of course be truly random, since each is determined by its predecessor in the sequence; but by the use of suitable generating equations it is possible to produce sequences in which the distribution of the numbers adequately satisfies the most important criteria of randomness, and the numbers themselves can reasonably be used as though they were random numbers. The generation of such pseudorandom number sequences is discussed by Lehmer (1951). By various devices, such as superimposing pseudorandomness on pseudorandomness, the approach to true randomness may be made even closer.

For the calculations described here, the Pseudorandom Number subroutine of the 'Pegasus' library was used. This uses the congruence $x_{n+1} \equiv kx_n \pmod{2^{31}-1}$, where x_{n+1} is the smallest positive integer satisfying this congruence, and k is the smallest positive integer satisfying the congruence $k \equiv 13^{13} \pmod{2^{31}-1}$. This produces in a pseudorandom sequence all numbers from 1 to $2^{31}-2$ inclusive; there are thus $2^{31}-2$ steps before the sequence is repeated. (Each 250-years run in the 37-population series uses about 2^{16} pseudorandom numbers; if the sequence is maintained, some 30 000 runs could be made before the initial value of x is met again.)

Each pseudorandom number produced consists of thirty-one binary digits, but before use here it is broken up into four sections of 3, 12, 3 and 12 digits, respectively, starting from the least significant end. The most significant digit is not used.

When the frequencies p_1, q_1, r_1 and s_1 have been calculated, they are uniformly scaled to integers P, Q, R and S, such that $P+Q+R+S = 2^{12}$. A pseudorandom number is generated, and the first of the 12-digit sections is extracted; this is the operative number, and is treated as a 12-digit integer, which may be any integer from 0 to $2^{12}-1$, inclusive, with equal probability. P, Q, R and S are then subtracted from this number in turn until

it becomes negative, when it is considered that a plant of the genotype corresponding to the last number subtracted has been taken into the sample; the chance of any genotype being sampled thus depends on the size of the relevant subtracted number, and is thus equal to its frequency as calculated, so long as the operative numbers are sufficiently random.

The approximation to randomness is improved by changing in a pseudorandom manner the order in which P, Q, R and S are subtracted. This is done by using the 3-digit section to choose one of eight predetermined arrangements of these four numbers before subtraction takes place. Since the eight possible 3-digit numbers will occur with equal frequency with each 12-digit number over the whole pseudorandom cycle, these two numbers are effectively independent of each other, and so each plant is sampled through two superimposed pseudorandom processes. When one plant has been sampled, the remaining 3- and 12-digit sections are used in the same way for a second plant. A new pseudorandom number is then calculated, and used as before for the third and fourth plants; and so on.

Of the twenty-four possible arrangements for subtraction only eight were used; more would have required extra digits and each generated number could have been used only once; this would have increased the computing time, while further gain in randomness would probably have been negligible. The eight arrangements were *PQRS*, *PRSQ*, *QSRP*, *QPSR*, *RQPS*, *RSQP*, *SRPQ* and *SPQR*.

This routine was tested independently of the evolutionary calculations by considering three quite different sets of values for P, Q, R and S, and two different sample sizes (64 and 32). For each combination in turn, a set of sixty-four samples was produced and its heterogeneity χ^2 determined. This was repeated, keeping the same order of the six frequency-sample-size combinations, until thirty-three sets of sixty-four samples had been obtained. The pseudorandom sequence was maintained during each set of samples, but a new sequence was started with each set; for sixteen of the sets, the pseudorandom number sequence fell within the large sequence used in producing the results described in this paper; the remaining seventeen were started at random.

The heterogeneity χ^2 for each set of samples has 189 degrees of freedom (n). Its mean value for the thirty-three sets was 186·4, with extreme values of 153·6, 158·3, 164·5, 218·9, 228·8 and 243·9. $\sqrt{(2\chi^2)}-\sqrt{(2n-1)}$ had a mean value of $-0{\cdot}13$ and a variance of 0·93; the expected values for a truly random process would be zero and 1·00, respectively; the standard error of this mean was $\pm 0{\cdot}168$. This indicates a very satisfactory simulation of a truly random sampling process.

Randomization of gene flow

The superimposing of random fluctuations on gene flow is more difficult to carry out satisfactorily, because an accurate assessment of the variance to be expected is not possible. The method chosen is a rough but reasonable guess based on my general knowledge of the populations in the field.

Gene flow is imagined to take place by considering that for each population pollen is contributed by a proportion of the plants of neighbouring populations, this proportion being determined by the distance apart and the size of the recipient population. The variance involved is not simply the sampling variance of this proportion, but will be a good deal greater. It will arise largely from differences in the size of this proportion from

year to year, resulting from differences in insect activity. The deviation from expectation will thus tend to be the same for all genotypes in any one year, and the simplifying assumption has been made that this is the rule; the direction is determined by reference to a single digit of a pseudorandom number.

The unit of variation will be the flower rather than the plant; as a working approximation, quarter-plant units have been used.

Each year, the numbers of quarter-plants to be expected as external pollen contributors in each genotype are determined for all populations before any calculations are carried out. When each population is reached in the calculation cycle, then for each genotype separately, a pseudorandom number is used to pick out one from 256 numbers packed at random in four blocks of the main store of the computer. These numbers range from 0 to 32 in frequencies which form the positive half of a flattened distribution curve which has 0 as the mode. When a number has been picked, it is scaled according to the expected number of quarter-plants, and is then added to or subtracted from the expected number, as previously decided. The scaling is such that the extremes of deviation are $\pm 100\,\%$ for the smallest expected numbers and $\pm 50\,\%$ for the largest. The modified numbers so obtained are then added in to the pollen parents for the recipient population, for that year.

This method is obviously rather crude, but it does produce a gene flow which seems to resemble natural flow; until it is possible to form a more accurate estimate of the variance of natural flow, further refinement may not be very helpful. It is proposed to test other methods, and compare the results.

References

Bateman, A. J. 1947*a* *J. Genet.* **48**, 257–275.

Bateman, A. J. 1947*b* *Heredity*, **1**, 303–336.

Crosby, J. L. 1948 Ph.D. thesis, University of Cambridge.

Crosby, J. L. 1949 *Evolution*, **3**, 212–230.

Crosby, J. L. 1959 *Heredity*, **13**, 127–131.

Fraser, A. S. 1957 *Aust. J. Biol. Sci.* **10**, 484–499.

Lehmer, D. H. 1951 *Proceedings of a second symposium on large-scale digital calculating machinery*, pp. 141–146. Harvard University Press.

Williams, W. T. & Lambert, J. M. 1959 *J. Ecol.* **47**, 83–101.

Wright, S. 1940 In *The new systematics*, pp. 161–183. Oxford University Press. (Edited J. S. Huxley.)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES