GENE FLOW MEDIATED BY WIND-BORNE POLLEN

J. T. GLEAVES

Department of Botany, Liverpool University

SUMMARY

It is shown that an inverse power law adequately describes the dispersion of airborne pollen over short distances. Two consequences of such a model are that the range of gene flow is inversely proportional to the square root of the population density, and that it is dependent upon spatial pattern; the range of gene flow decreases as the average nearest neighbour distance decreases with or without concomitant population density increase.

1. INTRODUCTION

ALTHOUGH a considerable amount has been published on clines, or the balance between selection and gene flow, there are still conspicuous gaps in our knowledge of the extent, and understanding of the properties of gene flow in plant populations.

Gene flow has been measured directly by the use of marker genes in spaced plant " garden " populations (Bateman, 1947; Griffiths, 1950; Copeland and Hardin, 1970), and in sward populations (Copeland and Hardin, 1970); and by measuring pollinator flight distances (Levin and Kerster, 1969). These measurements show that, although pollen can be blown hundreds of miles, and some insect pollinators (*e.g. Apis melifera*) frequently fly several miles, the bulk of cross-pollination occurs over very short distances.

Gene flow may be down to about 5 per cent. within 10 metres in spaced plant populations (Griffiths, 1950). Copeland and Hardin (1970), looking at genetic contamination in continuous swards in seed grower fields of rye-grass (using fluorescence as a marker), found that gene flow was negligible 6 metres from the border, and undetectable beyond 12 metres.

Gene flow has been estimated in natural populations, using those situations in which two genetically distinct populations adjoin and there is natural gene flow between them (parapatric populations), by comparing offspring with their parents, either with respect to physiological characters (such as tolerance to heavy metals) or with respect to morphological characters (Aston and Bradshaw, 1966; McNeilly, 1968; Watson, 1969; Snaydon, 1970). The effects of pollen dispersal have been shown to be remarkably limited, except in situations where there is a marked polarisation of wind direction (such as at Drws y Coed, where McNeilly found considerable gene flow over 150 metres).

Compared to our knowledge of the general extent of gene flow between such populations, relatively little is known about the form of the curve which describes the way in which the probability of two plants crossing decreases with distance. Bateman (1950), analysing gene flow data, has demonstrated that the curve is probably strongly leptokurtic. Levin and Kerster (1969)

have shown that bee mediated pollen dispersal (and therefore gene flow) is density dependent, and strongly leptokurtic.

Apart from the work of Bateman and of Levin and Kerster, I can find no theoretical treatment of gene flow except in conjunction with drift (Wright, 1946; Malecot, 1962), or with selection.

Where there is interaction between selection and gene flow, gene frequency clines result. Models of such clines fall into two categories.

(i) Mathematical models in which a function is sought which describes the change in gene frequency with respect to distance. The method used to find such a function is usually to set up an integral equation which represents an equilibrium in which changes in gene frequency due to migration are exactly balanced by changes due to selection, for every point in a continuous population. Provided that the gene-frequency/distance function (and its first derivative) is continuous with respect to distance, the integral equation can be approximated by a differential equation which can then be solved analytically (Haldane, 1948), or numerically (Fisher, 1950; Hansen, 1966), for different fitness régimes.

Because of the differential approximation, the probability density function, $f(s)$, which describes the distribution of migration distances, s, need never be stated explicitly. It need only be assumed to fulfil certain mathematically expedient requirements, such as symmetry about zero (odd moments zero), finity of all moments, and continuity (one such distribution is the normal distribution); it need only be characterised by its second moment (the mean square of the dispersal distance), giving rise to the " diffusion approximation ". Slatkin (1973) has shown that the " diffusion approximation " is only valid for weak selection pressures. In cases of strong selection, the detailed shape of $f(s)$ becomes important and its higher moments must be considered.

Downham (1973) has found direct numerical solutions to the integral equation, using a normal distribution for $f(s)$. He has shown that discontinuities with respect to distance in the relative fitness of the genotypes may give rise to discontinuities in the gene-frequency/distance relationship. These discontinuities are smoothed out by the " diffusion approximation ".

(ii) Numerical (computer simulation) models in which the interaction between different gene dispersal functions and different dominance/ fitness régimes is investigated. Jain and Bradshaw (1966) used the results of simulation in an attempt to decide whether, with different dispersal functions, the extremely steep clines found between para-patric plant populations, e.g. at the boundary of heavy metal mine waste tips, could be expected. They concluded that with lepto-kurtic pollen dispersal, steep clines were possible. This finding does not conflict with that of Bateman (1950).

My own approach to the theory of gene frequency clines is to look at the nature of gene flow, in the absence of selection, as a first step towards a fairly realistic model. In this paper, I shall present a model of gene flow and some of its implications for natural populations.

2. The model

In the simplest " population " of two plants, the amount of pollen each receives from the other can be expected to depend upon the distance between them (*i.e.* $p = f(x)$, where p is the amount of pollen received, and x is the distance between the plants). One can easily expand this idea to a population consisting of many plants with several genotypes. The amount of pollen received by a plant (A) from each genotype is the sum of the values of $f(x)$ corresponding to the distances between the plant (A) and each plant of the genotype in question. It should be observed that it is the proportion of each type of pollen which determines the genotypic frequencies among the progeny, therefore the function $f(x)$ need only describe relative amounts of pollen, not absolute amounts.

The model may be defined more formally as follows.

In a population of m plants containing w genotypes, the relative frequency of pollen from the uth genotype which is available to the ith plant is:

$$P_{i,u} = A_{i,u} / \sum_{v=1}^{w} A_{i,v}, \quad A_{i,u} = \sum_{j=1}^{n} \delta_{j,u} f(x_{i,j}),$$

$$\delta_{j,u} = \begin{cases} 1 \text{ if the } j\text{th plant is of genotype } u \\ 0 \text{ otherwise} \end{cases}$$

where $f(x_{i,j})$ is the amount of pollen received by the ith plant from the jth plant, $i \neq j$, and $f(x_{i,i}) = 0$ by definition.

In the computer, the position of each plant is defined by a pair of rectangular co-ordinates. Hence the distance between any pair of plants can be calculated by Pythagoras's theorem. For simplicity, each plant is assumed to produce the same amount of pollen, and all plants are regarded as being equally compatible to every other plant in the population, but totally self-sterile.

3. The choice of $f(x)$

The distribution of wind-blown pollen from its source (averaged over a flowering season) may be remarkably symmetrical (Griffiths, 1950; Raynor and Ogden, 1965), or it may be considerably asymmetrical (McNeilly, 1968). For simplicity, I have chosen to model the symmetrical situation. The most detailed observations on gene flow mediated by airborne pollen, which are suitable for use in fitting a model of pollen dispersal are those of Griffiths (1950).

The movements of airborne pollen may be described as eddy diffusion with deposition. Gregory's (1945) general formula for this type of movement has been expressed in simpler terms by Bateman (1947) as:

$$f(x) = x^{-(1+\frac{1}{2}d)} \exp \left\{ - \alpha x^{1-\frac{1}{2}d} / (1 - \tfrac{1}{2}d) \right\} \tag{1}$$

where x is distance, α is a constant, and d is a measure of turbulence.

Since it is the most appropriate formula found, I have attempted to fit it to Griffiths's results from experiment 3, 1948 (Griffiths 1950) (see fig. 1). It also seemed worth while to use an empirical formula:

$$f(x) = x^{-k} \exp \left(- \alpha x^{c} \right) \tag{2}$$

where x is distance, α, c, and k are constants. Bateman's formula (1) is a

special case of (2), as are the solutions to standard one, two, and three-dimensional diffusion equations, and the continuous functions used by Jain and Bradshaw (1966).

Fig. 1 shows the layout and sampling pattern of Griffiths's experiment no. 3. Due to the symmetry of the model and the experiment, it was only necessary to obtain simulation results for 24 plants in each of three rows in
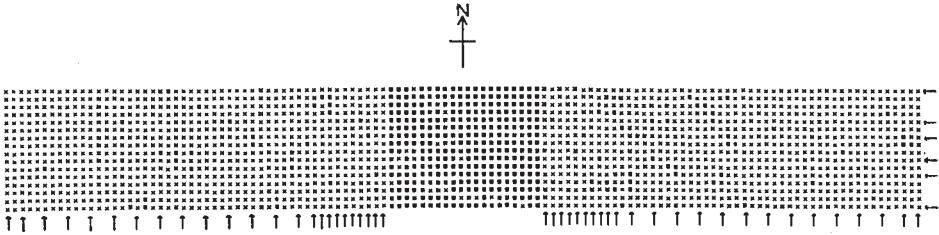

Fig. 1.—Griffiths's experiment 3, 1948 (Griffiths, 1950). X denotes a white-based plant, homozygous for a recessive allele c at one of the loci concerned with anthocyanin production. * denotes a red-based plant, homozygous for the dominant allelomorph C. The arrows mark the rows and columns sampled. The seed from the plants at the intersection of the marked rows and columns was germinated, and the frequency of the red-based phenotype scored. Distances are measured from the central block of red-based plants.

one arm of the experiment. Computation time was cut down still further by making use of the regularity of spacing.

The model was fitted by a numerical maximisation of

$$L = n\Sigma[p_o \ln (p_e) + (1-p_o) \ln (1-p_e)],$$

the logarithm of the likelihood function; n is the number of seeds counted per datum point ($\sim 400 \times 12 = 4800$, Griffiths, personal communication), p_o is the observed frequency of red-based seedlings, and p_e is that expected under the model.

Neither of the formulae (1) or (2) would give best fit with $\alpha > 0$. Negative values of α lead to $f(x)$ going through a minimum as x increases from zero, and then increasing as $x \rightarrow \infty$. The best realistic fit was given with $\alpha = 0$, $d = 1\cdot6694$ in formula (1), which then reduces to $f(x) = 1/x^k$,

$k = 1\cdot8347$. The standard error of k $\left(\text{estimated by } \sqrt{1/I}, I = \dfrac{-d^2L}{dk^2}\right)$ was

$0\cdot00627$. As can be seen in fig. 2, the fitted line does show the general trend among the points. However, the likelihood ratio test gives the highly significant value of chi-squared of $360\cdot063$, on 23 degrees of freedom (only one parameter estimated). It can be shown that no smooth curve of the type likely to be given by a pollen dispersal model will fit the data. In a crude and simple test, a smooth curve was fitted by eye in such a way that the trend (rather than any irregularity) was followed. A chi-square of $\sim 127\cdot5$, obtained from a likelihood ratio test, would be highly significant even if 23 degrees of freedom were allowed (*i.e.* even if only one " parameter " had been fitted in the smoothing of the data).

In view of the above results, it is suggested that in addition to the binomial sampling errors, there are other sources of variability about the expected

frequencies, two of which are fairly obvious. Firstly, there will almost certainly be differences between plants with respect to both number of florets and flowering time. If these differences are at least partly environmental in origin, they may be correlated with position. Secondly, a process
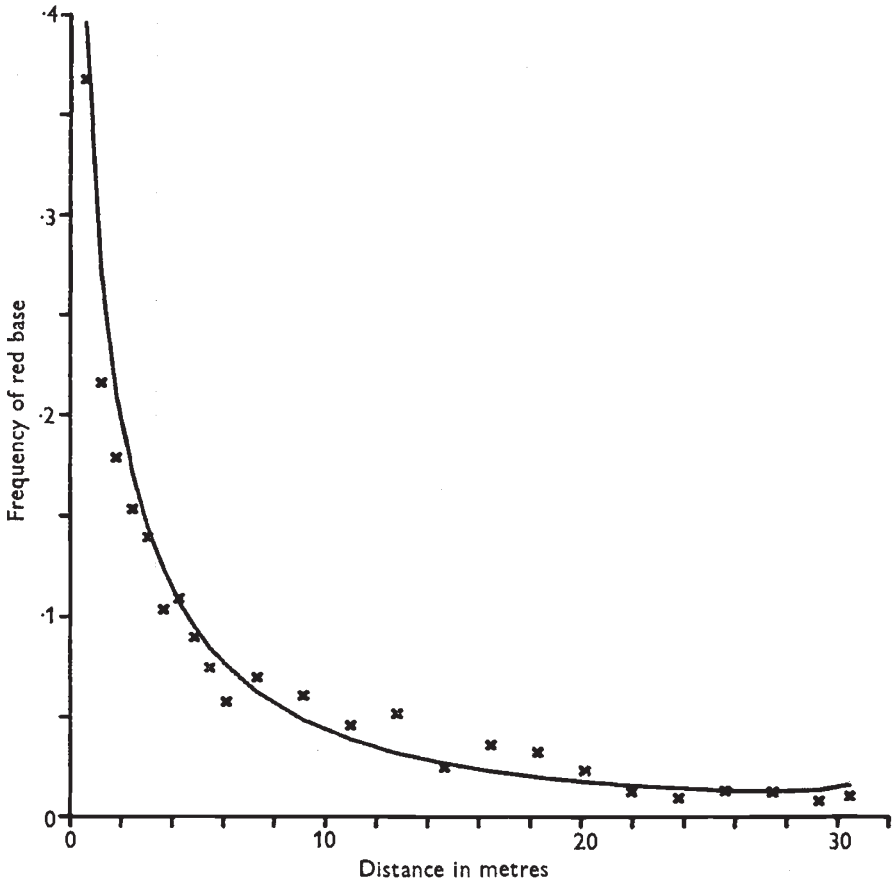

FIG. 2.—Frequency polygon of the best fit $(fx) = 1/x^{1\cdot8347}$) of the model of Griffiths's experiment 3. The data points (X) are the average of the east and west arms.

based on eddy diffusion cannot be expected to behave in a deterministic manner over a time period as short as the flowering season; the standard deviation of flowering time for *Lolium perenne* is 6·4 days (calculated from table 6, Griffiths, 1950).

Gregory (1945) maintains that deposition plays little part in determining the form of dispersion of airborne spores. Hence it may be argued that, if spores remain airborne for a considerable time, the mean square distance of dispersal may be very large (in comparison to the scale of Griffiths's experiment), and so the constant, α, of formula (1) (which is proportional to the deposition rate, and inversely proportional to the mean square dispersal distance) might be expected to be vanishingly small. Thus

$$\exp\{-\alpha x^{1-\frac{1}{2}d}/(1-\tfrac{1}{2}d)\} \simeq 1$$

over the distances considered. The value of $d = 1\cdot6694$ of formula (1) is within the range of $1\cdot24$-$2\cdot0$ contemplated by Sutton (1932) and Gregory (1945). Fig. 2 shows Griffiths's data (average of East and West arms) and the best fit of the model. It seems adequate for our purposes.

4. The effects of density

The constant k in $f(x) = 1/x^k$ is unitless. Hence in the simulation of any situation analogous to Griffiths's experiment, the units of distance are purely arbitrary; one achieves the same result whatever units of distance are used. The converse is also true; a graph of the relative frequency of red-based progeny against row number is unaffected by a change in the spacing of the

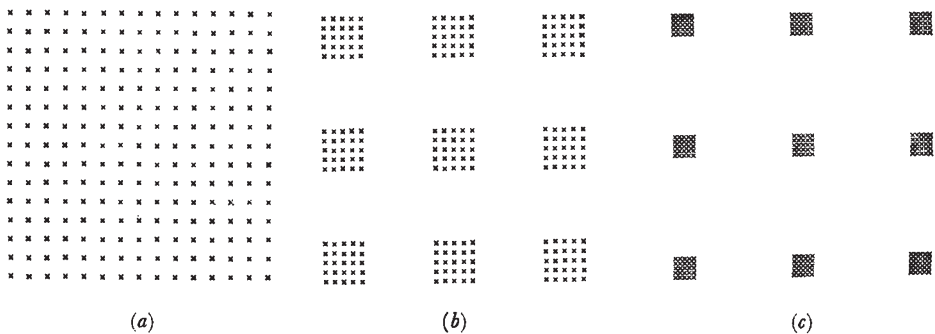(a)   (b)   (c)

Fig. 3.—280 cm. sided square populations of 225 spaced plants. (a) No aggregation, 20 cm. spacing. (b) 10 cm. spacing within aggregations. (c) 5 cm. spacing within aggregations.

plants (nearest neighbour distance). Therefore the 5 per cent. range of gene flow (the distance from the boundary within which gene flow is greater than 5 per cent.) is proportional to the nearest neighbour distance.

One might predict that, had Griffiths used 30·5 cm. spacing instead of 71 cm., he would have found a 5 per cent. range of gene flow of 5 metres instead of 10 metres. There is the proviso that the plants must be small with respect to the inter-plant distances; this is discussed later.

With regard to natural populations, one can infer from the above proposition that the range of gene flow (as defined above) is proportional to the average nearest neighbour distance. For a population in which the plants are uniformly distributed at random, the expectation of nearest neighbour distance (r) is given by:

$$E(r) = 1/(2\sqrt{\rho}),$$

where ρ is the population density (Clarke and Evans, 1954).

In natural populations, one might expect average nearest neighbour distance to be approximately inversely proportional to the square root of the population density.

Crude evidence for this density effect is to be found by comparison of Griffiths's spaced plant results with those of Copeland and Hardin (1970) from continuous swards. Both used species of *Lolium*. At distances beyond which gene flow was undetectable in continuous swards (12 metres), Griffiths was still getting 4 or 5 per cent. red-based progeny.

5. THE EFFECTS OF AGGREGATION

Random and regular distributions of plants are rare (Greig-Smith, 1964), but most models of clines apply strictly only to these. I have therefore investigated the effect on the amount of gene flow, of imposing aggregation on a regular or random distribution.

The model populations (fig. 3, *a*, *b* and *c*) take the form of square areas containing plants of one genotype, with and without aggregation, on to which falls a background rain of exogenous pollen. The source of exogenous
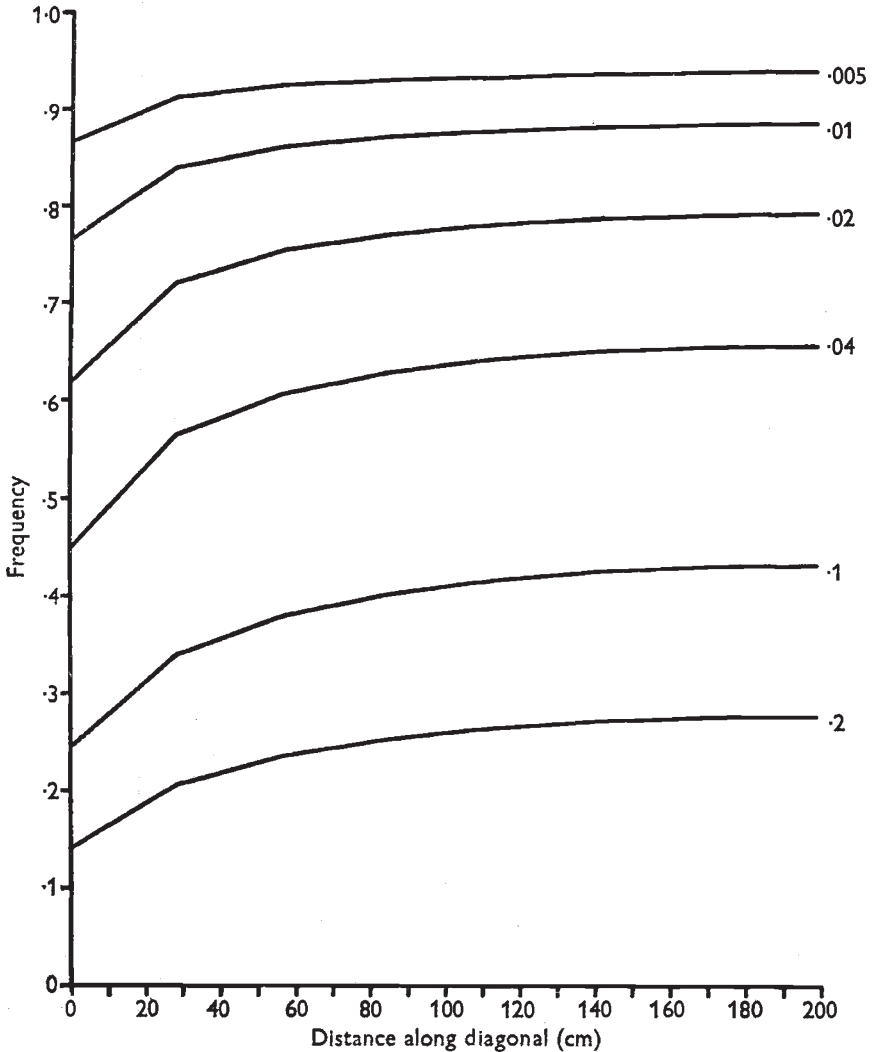
(*a*)

FIG. 4.—Frequencies of " within population fertilisation " for plants along the diagonal transect (from one corner at 0 cm. to the centre at 198 (280/√2) cm.) for different levels of background rain (0·005 to 0·2 relative units). Figs. *a*, *b* and *c* correspond to figs. 3*a*, 3*b* and 3*c* respectively.

31/3—Z 2

pollen is assumed to be at a considerable distance from the population, that is, each plant receives the same amount.   The amount of endogenous pollen received by each plant is the sum of all the values of $f(x)$ corresponding to the distances between each plant and every other plant in the population.
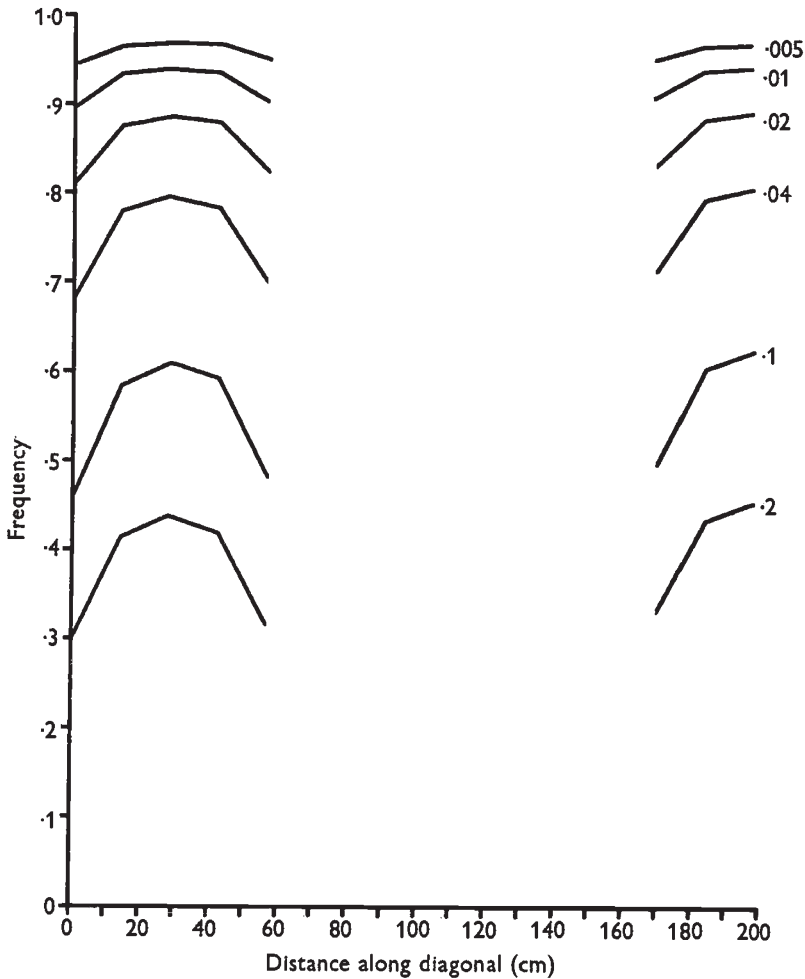

FIG. 4b

A relative unit of background rain is equal to $f(1)$, that is the relative amount of pollen that a plant would receive from another which was one unit of distance (*e.g.* one centimetre) away.

It can be seen from fig. 4, *a*, *b* and *c*, that more endogenous pollen is available to those plants at the centre of a group (either at the population, or the aggregation level), so that a smaller proportion of the pollen they receive is exogenous.   Fig. 5 summarises the effect of imposing aggregation onto a regular distribution of plants, both of the mean frequency of " within populations fertilisation ", and on the frequency along the diagonal transects. Fig. 6, which shows the effect of increasing the number of plants within a

square area, has been included for comparison. The effect of halving the interplant distances within an aggregation (containing a constant number of plants) is approximately the same as that of doubling the number of plants in the population.
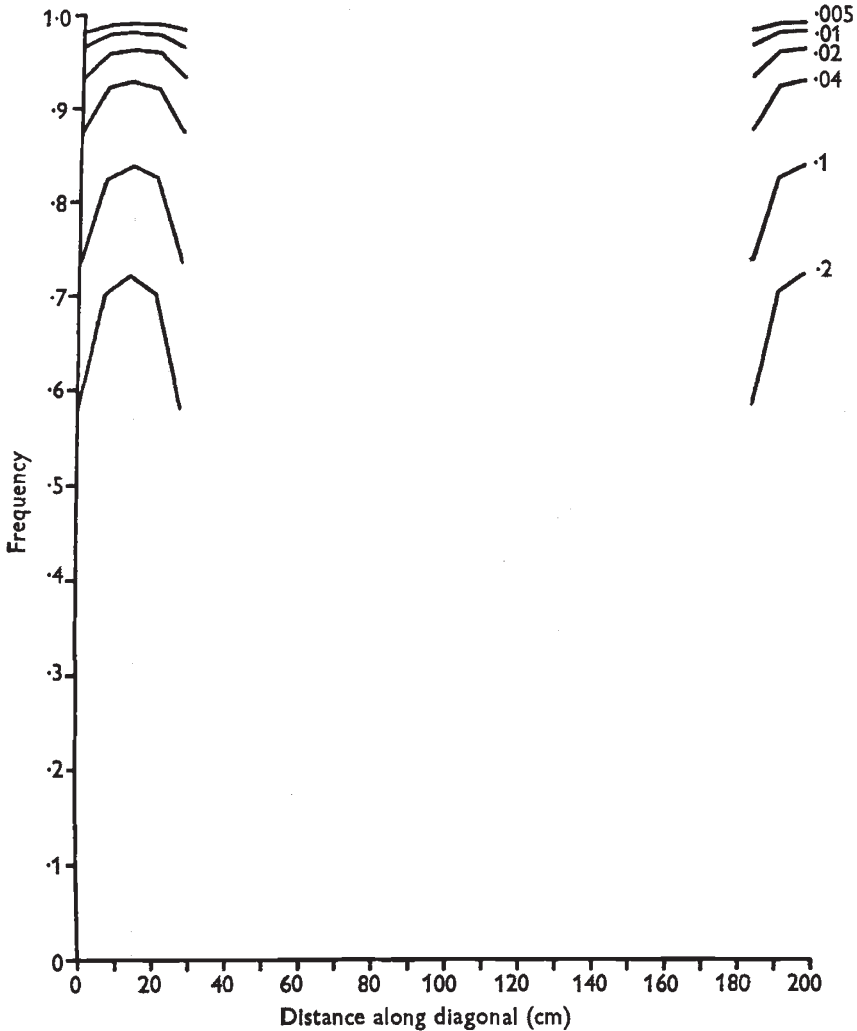

Fig. 4c

Results from simulation have shown that there is less gene flow into populations in which the plants are randomly distributed than into those in which plants are regularly spaced. With a background rain of 0·04 relative units, the average frequency of " within populations fertilisation " in square populations (225 plants in an area of 280 × 280 sq. cm.) was 0·602 for a regularly spaced population (fig. 3, a), and about 0·70 for randomly distributed populations. This result is not unexpected, as the average nearest neighbour distance (in populations of fixed density) is less for one with a random distribution than it is for one with a regular distribution (Clark and

Evans, 1954). If clumping is imposed on a random distribution, there is still less gene flow than into populations in which clumping is absent.
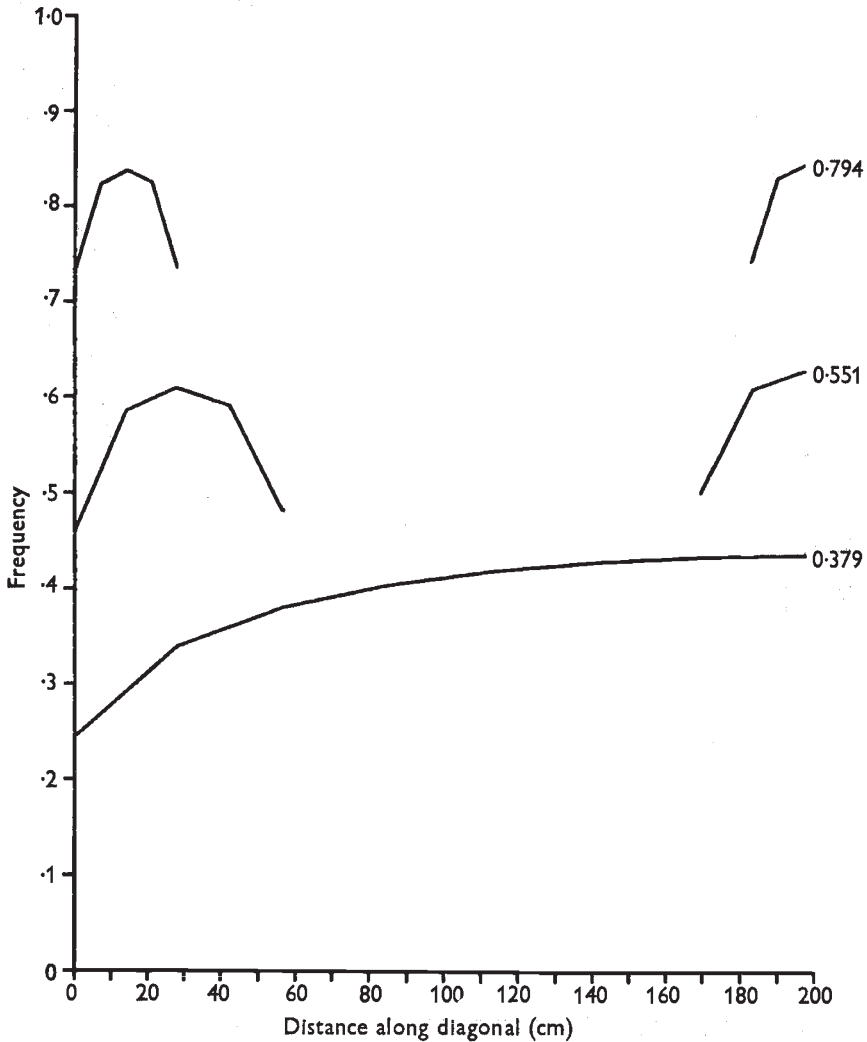

FIG. 5.—Comparison of " within population fertilisation " frequencies with different degrees of aggregation, for a background rain of 0·1 relative units. The numbers on the right-hand side of the lines are the averages (over all 225 plants) of the " within population fertilisation" frequencies.

In general, gene flow into an area decreases as the average nearest neighbour distance (within that area) decreases, whether this be due to an increase in clumping, or to a general increase in population density.

6. DISCUSSION

One of the less realistic features of the model presented here is the representation of plants by points in two-dimensional space. The effect of

finite plant size is to decrease the effective inter-plant distances from x to $x-b$, where b is a function of plant size, and x is the distance between the centres of two plants. The shorter the inter-plant distance, the greater the effect of plant size on the amount of pollen flow between two plants. Therefore, representation of plants as points will tend to overestimate the range of
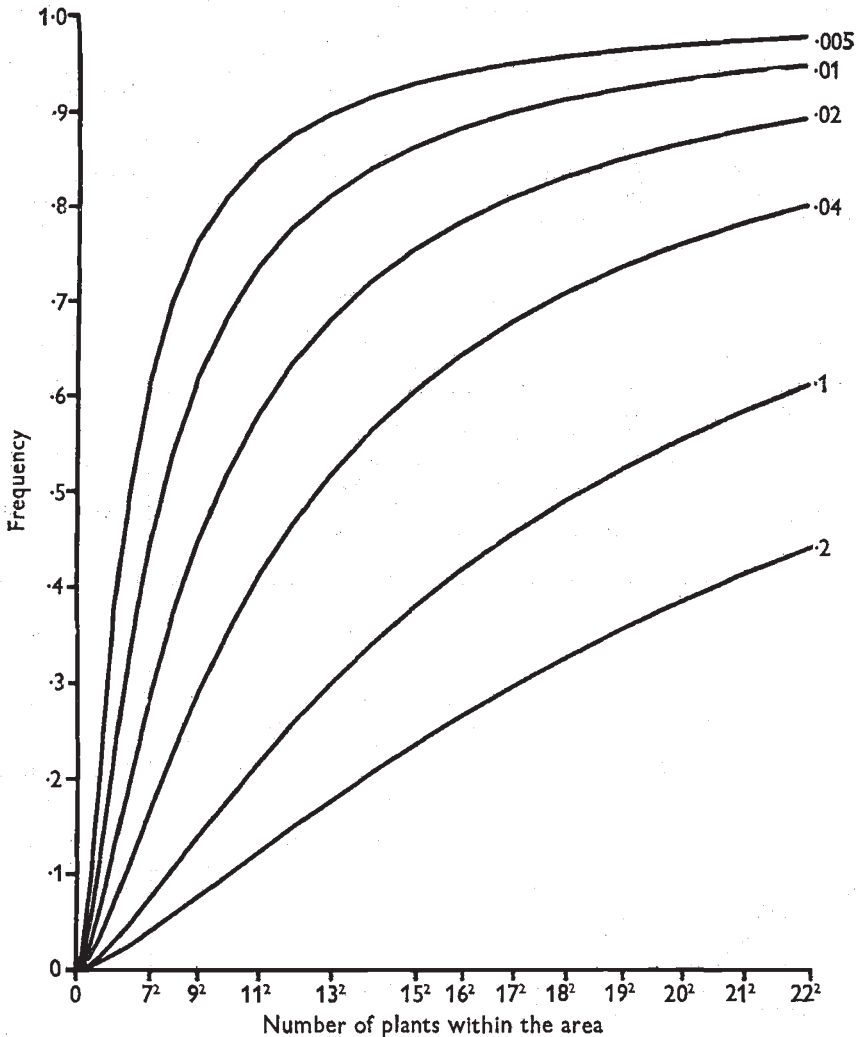

FIG. 6.—The effect on the average " within population fertilisation " frequency of increasing the number of plants within a regularly spaced, 280 cm. sided, square population, at different levels (0·005 to 0·2 relative units) of background rain.

gene flow, and the larger are the plants in proportion to the inter-plant distances, the greater the overestimate will be. In consequence, the effects of density and aggregation may have been underestimated. However, the situation is further complicated because plant size may be affected by the competitive effects arising from nearest neighbour distance.

Jain and Bradshaw (1966) found that in some cases, clines were steeper than would be suggested theoretically. They felt that, either selection must be stronger than estimated, or that gene flow was restricted in ways not appreciated. It should be noted that they based their range of gene flow on that reported by Griffiths (1950), assuming that this range was not dependent upon population density or pattern. In fact, Griffiths used regularly spaced plants at very low density. In the natural situations which Jain and Bradshaw examined, densities are likely to be much higher, and distributions certainly not regular. In the light of the present investigation, one might expect gene flow to be more restricted than previously envisaged.

7. REFERENCES

ASTON, J. L., AND BRADSHAW, A. D. 1966. Evolution in closely adjacent plant populations. II. *Agrostis stolonifera* in maritime habitats. *Heredity, 21*, 649-664.

BATEMAN, A. J. 1947. Contamination in seed crops. III. Relation with isolation distance. *Heredity, 1*, 303-336.

BATEMAN, A. J. 1950. Is gene dispersal normal? *Heredity, 4*, 353-363.

CLARK, P. J., AND EVANS, F. C. 1954. Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology, 35*, 445-453.

COPELAND, L. O., AND HARDIN, E. E. 1970. Outcrossing in ryegrass (*Lolium spp.*) as determined by fluorescence tests. *Crop Science, 10*, 254-257.

DOWNHAM, D. Y. 1973. Personal communication.

FISHER, R. A. 1950. Gene frequencies in a cline determined by selection and diffusion. *Biometrics, 6*, 353-361.

GREGORY, P. H. 1945. The dispersion of air-borne spores. *Trans. Brit. Mycol. Soc., 28*, 26-72.

GREIG-SMITH, P. 1964. *Quantitative Plant Ecology.* Butterworth.

GRIFFITHS, D. J. 1950. The liability of seed crops of perennial ryegrass (*Lolium perenne*) to contamination by wind borne pollen. *J. Agri. Sci., 40*, 19-38.

HALDANE, J. B. S. 1948. The theory of a cline. *J. Genet., 48*, 277-284.

HANSEN, W. D. 1966. Effects of partial isolation (distance), migration and different fitness requirements among environmental pockets upon steady state gene frequencies. *Biometrics, 22*, 453-468.

JAIN, S. K., AND BRADSHAW, A. D. 1966. Evolutionary divergence among adjacent plant populations. I. The evidence and its theoretical analysis. *Heredity, 21*, 407-441.

LEVIN, D. A., AND KERSTER, H. W. 1969. Density dependent gene dispersal in Liatris. *American Naturalist, 103*, 61-74.

MALECOT, G. 1962. Migration et paranté génétique moyenne. *Entretiens de Monaco en Sciences Humaines*, 205-212.

MCNEILLY, T. 1968. Evolution in closely adjacent plant populations. III. *Agrostis tenuis* on a small copper mine. *Heredity, 23*, 99-108.

RAYNOR, G. S., AND OGDEN, E. C. 1965. Twenty-four hour dispersion of Ragweed pollen from known sources. Brookhaven National Laboratory, BNL 957 (T-398), Upton, N.Y.

SLATKIN, M. 1973. *Genetics*, in the press.

SNAYDON, R. W. 1970. Rapid population differentiation in a mosaic environment. I. The response of *Anthoxanthum odoratum* populations to soils. *Evolution, 24*, 257-269.

SUTTON, O. G. 1932. A theory of Eddy diffusion in the atmosphere. *Proc. Roy. Soc. A, 135*, 143-165.

WATSON, P. J. 1969. Evolution in closely adjacent plant populations. VI. An entomophilous species, *Potentilla erecta*, in two contrasting habitats. *Heredity, 24*, 407-422.

WRIGHT, S. 1946. Isolation by distance under diverse systems of mating. *Genetics, 31*, 39-59.