"THE GENETICAL STRUCTURE OF SUBDIVIDED POPULATIONS"

Thesis submitted for the degree of Doctor of Philosophy

in the University of Adelaide by

GREGORY CHARLES KIRBY B.Sc.(Hons.)

Department of Genetics

University of Adelaide

September 1974.

TABLE OF CONTENTS

SUMMARY

Mathematical models of population genetics normally assume panmixia. But natural populations often depart from panmixia because of barriers to gene flow, caused by such factors as geographic variation in habitats, behaviour patterns that discourage immigration, or isolation by distance. To deal with these situations, various mathematical theories about subdivided populations have been developed. The validity of predictions derived from such mathematical theories should be tested against data collected from carefully chosen biological populations. This thesis reviews the mathematical models of subdivided populations, presents additional results and provides an analysis of data collected from two studies on populations of wild house mice (*Mus musculus*).

Chapter 2 reviews the models used to study the genetical properties of subdivided populations and discusses the results obtained from these models.

The theory of F_{IS} (the fixation index within subpopulations) is developed for a dioecious population, taking account of the adult stage of the lifecycle (Chapter 3). Two definitions of F_{IS} (due to Wright and Nei) are compared and Wright's definition is shown to result in simpler formulae. The expected value of F_{IS} is shown to be negative in an isolated subpopulation and in a subpopulation receiving immigrants. The theory of F_{IS} is applied to data from natural populations.

An estimator of the population fixation index, given a sample fixation index and sample size, is derived in Chapter 4. Computer simulation shows this estimator is apparently unbiassed.

Chapter 5 considers the relationship between the F-statistics of Wright and two measures of the probability of identity of two randomly chosen homologous genes.

Studies on the ecological genetics of the house mouse are reviewed in Chapter 6 where it is concluded that this species is eminently suitable for studying the effects of population subdivision in nature. The remainder of this thesis is concerned with such an investigation.

The methods used to score eight loci controlling electrophoretic variation of plasma and red cell proteins and esterases in mice are briefly described (Chapter 7).

The results of crosses set up to check the inheritance of the phenotypes produced by eight previously described loci are given in Chapter 8. The phenotypes at four of these loci were affected by non-genetic factors, so the data collected for these loci was discarded. An account is given of the inheritance of variants at two loci thought to be previously undescribed.

Chapter 9 presents an investigation of the genetical structure of four different mouse populations. This had two objectives, firstly, a simple demonstration that different ecological conditions beget different genetic structures and secondly, the testing of three theoretical predictions against real data. These predictions are:

(1) When a single subpopulation is sampled, the expected fixation index is negative, but if many subpopulations are included in a sample, it is positive;

(2) the amount of genetic differentiation between subpopulations within a population is inversely proportional to Nm, where N is the mean effective size of subpopulations and m is the mean effective rate of immigration into subpopulations;

(3) the genetic divergence between subpopulations should increase as the geographic distance between them increases.

The data obtained show that the genetical structure of a population changes if the ecological conditions alter. Predictions (1) and (2) are in agreement with the data, but for populations in which the average

geographic distance between subpopulations is less than 120km, there is no correlation between genetic and geographic distances.

Chapter 10 describes a trap-bleed-release-recapture study carried out on three small mouse subpopulations on farmlands. Each subpopulation was supplied with surplus food from three to six feedstations (under which the mice lived in burrows) and demographic data indicate that the mean effective breeding number was only about two per feedstation. Each mouse caught was scored for five loci. Marked gene frequency changes occurred in these small subpopulations due to random drift and immigration. Genotype frequencies in these subpopulations departed significantly from Hardy-Weinberg frequencies. The theory of F_{IS} developed earlier shows that the observed excess of heterozygotes could be attributed to the small size of the subpopulations. An attempt was made to estimate the rate of immigration into each subpopulation and the local equilibrium gene frequencies from the regression of Δq on q, following Tamarin and Krebs. However, it is shown in an appendix that the estimates obtained from this regression are biassed.

A short note on the relation between the effective and actual numbers of alleles in samples from natural populations is given in an appendix.

Declaration

I declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any University and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due reference is made.

G. C. Kirby

ACKNOWLEDGEMENTS

I am particularly grateful to my supervisor, Professor J.H. Bennett, for his encouragement, tolerance and sagacious advice. Dr. A.E. Newsome encouraged me to work on the house-mice at Turretfield and he willingly loaned to me his mouse traps and unpublished data. Dr. R.M. Hope taught me the art of starch gel electrophoresis. Mr. C. Chesson provided invaluable assistance and advice about the maintenance of mice while other technical assistance was given by Mr. I.R. Goodwins and Mr. J.R.B. Walpole. Dr. G.M.E. Mayo provided statistical advice.

I must thank the following for their patience and assistance in checking the final drafts of this thesis: Dr. G.M.E. Mayo; Dr. O. Mayo; Dr. B.W. Gabb; Mr. J. Oakeshott and Mr. R.B. Halliday.

Dr. D.L. Hayman also helped with the work required for this thesis.

Many people assisted me to catch mice: Mr. I.R. Goodwins, Dr. R.H. Maddern, Mr. J. Oakeshott, Mr. I. Maynard, Mr. P. Close, Mr. L.H. Schmitt Dr. P. Reeves and Mr. P. Aitken.

The following persons or organisations kindly allowed me to catch mice from land or buildings owned or controlled by them: Mr. H. Day (Turretfield Agricultural Research Centre); Mr. C. Wheatcroft (Price); Mr. D. Hearn (South Kilkerran); Mr. B. Snodgrass (Alford); Mrs. Leighwood (Port Vincent); Mr. P. Watters (Minlaton); Mr. G. Price (Wardang Island); Messrs. Tsemelis, Oswalds and Wiscesko of Port Lincoln; Messrs. Lancaster and Werner of the Adelaide Zoological Gardens; Mr. Gates and the C.S.I.R.O. Nutritional Biochemistry Division of O'Halloran Hill; Key's Pet Shop of Adelaide; and Mr. A. McNeil of the Waite Agricultural Research Institute.

The typing was done by Miss G. Psaltis and Mrs. A. McKay (who was a source of inspiration with her unflagging enthusiasm and encouragement).

My wife's support made the completion of this thesis possible.

I am grateful to the University of Adelaide for financial support, firstly on a University Research Grant Scholarship (1969-1973) and later on a Demonstratorship (1973-1974).

CHAPTER 1

INTRODUCTION

Most mathematical models of population genetics assume a panmictic population and an absence of geographically localised selective values. These two assumptions simplify equations which are intended to show, for example, the effects of linkage, selection or different breeding systems on gene and genotype frequencies. But attempts to predict what should occur, or to interpret what has occurred, in populations that are subdivided, will receive little assistance from such simplified models. Population subdivision occurs because of barriers to gene flow (e.g. physical barriers such as inhospitable territory between subpopulations or isolation by distance, and behavioural barriers such as territoriality) or because of environmental heterogeneity that changes the selective value of a gene from place to place. Most domestic animals and plants also have subdivided populations because each species is broken up into breeds or strains which are usually not allowed to interbreed. This thesis will emphasise natural rather than domestic populations, although much of the theory applies to both situations.

The study of subdivided populations has two main objectives. Firstly, to analyse the possible causes (and consequences) of genetic differences between localities. Secondly, to develop methods of analysis that can identify which factor or factors caused the observed genetic differentiation of subpopulations in a particular population. Several extra parameters are needed for mathematical models of subdivided and genetically differentiated populations. Because gene frequencies vary between subpopulations, the population gene frequency must be supplemented by a measure of the variation in gene frequencies across the population (the variance in gene frequencies or an inbreeding coefficient, are often

used). The differentiation of gene frequencies is usually controlled by a balance between three factors: random drift due to the small size of subpopulations; selection; and gene exchange between subpopulations. In the mathematical study of these factors it is often difficult to obtain simple equations that express the relationships between these factors and other parameters, but these difficulties must be surmounted if the evolutionary processes within subdivided populations are to be understood.

The study of the existing mathematical theory of subdivided populations is impeded by the diversity of methods that have been used to deal with the problems of population subdivision. Rarely are the different methods compared and often the similarities and differences between them are difficult to identify. In an attempt to rectify this situation, a relationship between two methods (F-statistics and probabilities of identity) of measuring the amount of genetic variation within and between subdivisions is given in Chapter 6.

A few examples should serve to illustrate the effects of population subdivision in natural populations. The human races not only show great differences in conspicuous features such as skin colour and hair texture, but also differ in gene frequencies for such hidden characters as blood groups (e.g. Cavalli-Sforza *et al.*, 1964). In several species of moths, the frequency of melanic moths change along gradients in pollution levels (e.g. *Biston betularia;* Bishop 1972) or climate (e.g. *Amathes glareosa;* Kettlewell and Berry 1961). Snails can have marked changes in the frequencies of shell colours and banding patterns from one colony to the next, which sometimes correlate with environmental changes and sometimes do not (e.g. *Cepaea nemoralis;* Cain and Currey 1963; Arnold 1969). Extensive genetic variation between neighboring plant populations is found for several characters in different species (e.g. Jain and Bradshaw, 1966).

Recent advances in the techniques for studying genetical variation in natural and mathematical populations have increased the number of genetical studies of natural populations and given fresh insight into the amount of variation within and between subpopulations in a species. Firstly, the technique of electrophoresis, combined with appropriate staining techniques, allows the detection of many loci in most species (from bacteria (Milkman, 1973) to humans (Harris, *et al.*,1974)). Often these loci have codominantly inherited alleles that control electrophoretic variants. Secondly, computers can be used to evaluate (by the solution of equations or simulation) mathematical models of increasing genetical and ecological complexity, from the simple simulation of random genetic drift (e.g. Rohlf and Schnell, 1971) to the conditions for sympatric speciation (e.g. James, 1970; Dickenson and Antonovics, 1973). Thirdly, genetic distance statistics have been devised to estimate the divergence (or similarity) between subpopulations in which gene frequencies have been estimated for several loci (e.g. Edwards, 1971; Hedrick, 1971; Nei, 1972; Latter, 1973). Other multivariate statistical methods are also being introduced (e.g. Gould and Johnston, 1972).

Many of the recently discovered isozyme polymorphisms show geographic variation in gene frequencies, but possible reasons for the variation are only rarely known. In some cases, evidence has been found to suggest that the selection pressures acting on the alleles at a locus are changing along an environmental gradient (for example, a temperature gradient has been implicated in the following cases: O'Gower and Nicol, 1968; Koehn, 1969; Schopf and Gooch, 1971; Johnson, 1971).

In many studies of genetic variation in natural populations, the results are presented in extensive tables of gene frequencies with a little discussion about the variation within and between samples and the possibilities for selection, random drift and gene flow. But, far

too often, no useful data about the ecology of the species are included (e.g. population densities, dispersal distances, age distribution) and without such data, the contemporary arguments about selection, random drift and gene flow will be difficult to resolve (Lewontin, 1967). Rarely are theoretical models and field observations of subdivided populations put together so that the theory and the observations complement each other.

One of my intentions in this thesis is to relate mathematical theory to observations and vice versa. The mathematical work on the frequency of heterozygotes in small subpopulations and on the bias in the regression of Δq on q, grew out of the need to interpret certain results from field studies. A comparative study of genetical variation between subpopulations within four different populations was carried out as a test of the simple theory that the amount of genetical variation between subpopulations is inversely proportional to subpopulation sizes and immigration rates. These data were also used to test two further hypotheses about the genetical structure of subdivided populations. A second study of natural populations was intended to compare independent ecological and genetical estimates of the effective subpopulation sizes and immigration rates in three subpopulations.

The house mouse (*Mus musculus*) was chosen as a suitable organism for field studies on a subdivided population because the ecology and genetics of this species are well known.

CHAPTER 2

MATHEMATICAL THEORY OF SUBDIVIDED POPULATIONS

2.1 BASIC THEORY OF A PANMICTIC POPULATION

2.1.1 Hardy-Weinberg Law

The Hardy-Weinberg law is used to find the expected genotype frequencies in a random mating diploid population. The multi-allelic generalisation of this law can be expressed as follows: if there are ℓ alleles (a_1, a_2, a_3 .... a_ℓ) at a locus with frequencies q_1, q_2, q_3 .... q_ℓ, respectively, and fully random mating occurs, the expected frequency of any homozygote, $a_i a_i$, is q_i^2 and the expected frequency of the $a_i a_j$ heterozygote is $2q_i q_j$.

2.1.2 Random Sampling Drift

Any real population contains only a finite number of individuals. The genes present in each successive generation are only a sample of those present in the previous generation and hence the frequency of a gene that is not fixed will alter from one generation to the next. This gene frequency change is undirected and its magnitude is a random variable. Cain and Currey (1963) and Wright (1970) have suggested that gene frequency changes due only to finite population number be known as random sampling drift.

If the gene frequency is q_t in generation t, then

$$q_{t+1} = q_t + \delta q_t$$

where δq_t is the random sampling drift in generation t. The important properties of δq_t are that its mean is zero (i.e. $E(\delta q_t) = 0$ where $E(\cdot)$ is the operation of mathematical expectation) and its variance is

$$E(\delta q_t)^2 = q_t(1-q_t)/2N \qquad \ldots\ldots(2.1)$$

where N is the number of diploids in the population.

The early studies on the importance of random sampling drift emphasised the role this process plays in reducing genetic variability in a population because genes can only drift until extinction or fixation (Fisher, 1922,1930; Haldane, 1927; Wright, 1929). Fisher and Haldane dealt in detail with the probability of extinction of a gene newly introduced into the population. Whilst only a few representatives of a gene are present in a population, the probability of extinction for the gene is high and depends more upon the gene frequency than the selective value that it possesses (Haldane, 1927; Fisher, 1930).

2.1.3 Mutation

Mutation involves a physical change in the genetic material and is the mechanism that generates new alleles at a locus or re-creates old alleles. Mutation has a simple effect on gene frequencies but acts only slowly because of the low rate at which mutations occur. If an allele with frequency q_t mutates at a rate u per generation and back-mutation occurs at rate v per generation, the change in gene frequency per generation is

$$\Delta q_t = -uq_t + v(1-q_t) \qquad \ldots\ldots(2.2)$$

Provided u,v >0 and mutation is the only factor changing gene frequencies, an equilibrium gene frequency ($\bar{q}$) will be attained:

$$\bar{q} = v/(u+v) \qquad \ldots\ldots(2.3)$$

and Eq.(2.2) can be rewritten

$$\Delta q_t = (u+v)(\bar{q} - q_t) \qquad \ldots\ldots(2.4)$$

However, modern knowledge of the structure of genes suggests that this simple model of mutational equilibrium may be inadequate (Kimura and Crow, 1964; King and Jukes, 1969; Kimura, 1968; Kimura and Ohta, 1971a,

1971b; Crow, 1970). Consider, for example, the haemoglobin α-chain (*Hbα*) gene in humans which has 432 nucleotides. A single point mutation can change any one of these nucleotides to a different one (adenine, thymine, guanine and cytosine), creating a new ($Hb\alpha^1$) allele. If this allele is not quickly lost by random sampling drift, it will also undergo mutation. But the odds on this next mutation occurring at the same site as before are only 1:431 and even if at the same site, the probability of returning to the same nucleotide as before is only 1/3. Thus mutations of the $Hb\alpha^1$ are much more likely to produce new alleles and only very rarely will back-mutations occur.

This line of argument led Kimura and Crow (1964) to suggest a model of mutational evolution in which every new mutant allele is different from any of the alleles pre-existing in the population. There is a flux of new alleles entering the population by mutation and being eliminated by random sampling drift in this "infinite alleles" model. The genetical state of the population can no longer be described in terms of equilibrium gene frequencies but must be described by the expected number of different alleles and the effective number of alleles present in the population when an equilibrium is established between mutational gain and random loss of alleles.

This model of evolution by the substitution of neutral or nearly neutral alleles has been championed by Kimura and Ohta (1971a,1971b), but many are opposed to it (e.g. Clarke, 1970; Ayala *et al.*,1972). The *Hbα* example illustrates two difficulties with the model. Firstly, the assumption that all 432 nucleotide sites in the *Hbα* gene can mutate freely to neutral alleles is erroneous because some of the amino acids in the molecule cannot be changed without destroying the function of the molecule (see Giblett, 1969, for review). Secondly, because of the synonymity of codons for amino acids, changes in many nucleotides will

not be detectable by conventional biochemical techniques since amino acid substitutions will not occur.

In practice, the range of detectable amino acid variations that are selectively neutral may be few in number, but at present there does not appear to be any crucial evidence on this point (compare Johnson, 1972; Kirby and Halliday, 1973 (Appendix 4); and Yamazaki and Maruyama, 1973; for three viewpoints on the same data).

2.1.4 Migration

Migration is the entry of genes from another population. If the gene frequency of immigrants is $\bar{q}$ and the immigrants are a proportion m of the breeding population in generation t,

$$q_{t+1} = (1-m)q_t + m\bar{q}$$

$$\Delta q_t = m(\bar{q}-q_t) \qquad \ldots\ldots (2.5)$$

Inspection of equations (2.4) and (2.5) reveals that migration and reversible mutation are mathematically equivalent. But there is an important biological difference between the two: mutation affects only one locus at a time, while migration affects gene frequencies at many loci simultaneously. The magnitude of m in natural populations is normally much greater than (u+v).

2.1.5 Selection

Wright (1969) defined selection as "... any process in a population that alters gene frequency in a directed fashion without change of the genetic material (mutation) or introduction from without (immigration)." Selection includes an enormous variety of processes that can occur at any stage of the life-cycle. Despite the bewildering array of possible ways for selection to occur, only a few situations have been regarded as sufficiently likely and reasonable to be dealt with in detail (Haldane,

1932; Fisher, 1930; Wright, 1931, 1969; Crow and Kimura, 1970).

Genic selection applies to haploid organisms where one allele is in the process of replacing another and to diploid organisms when the heterozygote is exactly halfway between the two homozygotes in relative fitness. In the haploid case, the relative fitnesses of alleles a_1 and a_2 are in the ratio (1-s):1 (s is the selection coefficient) whereas in the diploid case the relative fitnesses of a_1a_1, a_1a_2 and a_2a_2 are (1-2s): (1-s):1. The change in the frequency of a_1 (q_t) over one generation is approximately (ignoring a small term in the denominator if s is small),

$$\Delta q_t = -sq_t(1-q_t) \qquad \ldots\ldots(2.7)$$

Genotypic selection occurs when one allele is recessive to the other with respect to fitness. Let the relative fitnesses of the dominant phenotype and the recessive phenotype be 1:(1-s). The change in frequency of the recessive allele is, approximately,

$$\Delta q_t = -sq_t^2(1-q_t) \qquad \ldots\ldots(2.8)$$

A balanced polymorphism (Ford, 1964) can be established when the heterozygote always has greater fitness than either of the homozygotes (Fisher, 1922), or when the fitnesses of the genotypes are frequency dependent such that the average fitness of each allele (averaged over all genotypes) decreases as its frequency increases. In the first case (heterosis or overdominance) the relative fitnesses of the genotypes a_1a_1, a_1a_2 and a_2a_2 are (1-s):1:(1-r).

Approximately, $$\Delta q_t = q_t(1-q_t)(r-(s+r)q_t) \qquad \ldots\ldots(2.9)$$

and provided that $0<s,r<1$, these will be an equilibrium at

$$\bar{q} = r/(s+r) \qquad \ldots\ldots(2.10)$$

and equation (2.9) can be reformulated as

$$\Delta q_t = q_t(1-q_t)(s+r)(\bar{q}-q_t) \qquad \ldots\ldots(2.11)$$

Comparison of equations (2.4),(2.5) and (2.11) shows that heterosis is similar to recurrent mutation and immigration in its effect on gene frequencies, provided q_t is not too different from $\bar{q}$.

Wright (1969) and Cook (1971) reviewed several mathematical models for frequency dependent selection, but it is not clear which models, if any, will be relevant to natural populations. Cook (1971) reviewed circumstances when frequency dependent selection probably occurred: *Drosophila melanogaster* Esterase-6 locus (Kojima and Yarbrough,1967); selective predation; Batesian mimicry; and chromosome inversions in *Drosophila pseudoobscura* (Wright and Dobzhansky, 1946). Nassar *et al.* (1973) produced evidence for frequency dependent selection on the Payne inversion in *Drosophila melanogaster* under crowded conditions.

An equilibrium can also be established between selection and mutation to a deleterious gene. Ignoring back mutation from the deleterious gene, the combined effect of mutation and genic selection is, approximately,

$$\Delta q_t = v(1-q_t) - sq_t(1-q_t) \qquad \ldots\ldots(2.12)$$

The equilibrium gene frequency, $\bar{q}$, will be small;

$$\bar{q} = v/s \qquad \ldots\ldots(2.13)$$

For a recessive deleterious gene,

$$\Delta q_t = v(1-q_t) - sq_t^2(1-q_t) \qquad \ldots\ldots(2.14)$$

$$\bar{q} = \sqrt{v/s}$$

The frequency of strongly deleterious genes is expected to be negligible at any locus, but the cumulative frequency over the thousands of loci in the genome may be significant. Dobzhansky and Wright (1941) found that 15% of the third chromosomes in a *Drosophila pseudoobscura* population carried recessive lethal genes.

2.2 MODELS AND METHODS FOR THE THEORETICAL STUDY OF SUBDIVIDED POPULATIONS.

2.2.1 The Island Model

Wright (1921,1922) introduced the concept of the correlation between uniting gametes, which he denoted by F, for fixation index. F measures the deviation from Hardy-Weinberg frequencies and is also known as the inbreeding coefficient of the population. When there are two alleles with frequencies q and (1-q) and a fixation index F, the frequency of heterozygotes in the population (H) will be

$$H = 2q(1-q)(1-F) \qquad \ldots\ldots(2.16)$$

and hence

$$F = 1 - H/2q(1-q) \qquad \ldots\ldots(2.17)$$

Wright (1931) pointed out that population subdivision had the same effect on genotype frequencies as inbreeding. If σ_q^2 is the variance in gene frequencies between subdivisions and there are Hardy-Weinberg frequencies within subdivisions, the frequency of heterozygotes in the whole population is

$$H = 2q(1-q) - \sigma_q^2 \quad , \qquad \ldots\ldots(2.18)$$

which is a formula due to Wahlund. Inspection of equations (2.16) and (2.18) shows that in a subdivided population,

$$\sigma_q^2 = Fq(1-q)$$

or

$$F = \sigma_q^2/q(1-q) \qquad \ldots\ldots(2.19)$$

If σ_q^2 is known, then F can be calculated and vice versa.

The first major contribution to the mathematical theory of subdivided populations was Wright (1931) in which was introduced the "island model" of population structure. This model assumes a population subdivided into islands within which there is random breeding and between which migration occurs at random, regardless of the distance between islands. Each island consists of N diploids and a proportion, m, of

these are replaced by immigrants each generation. In the simple version of the model, it is implicitly assumed that there are an infinte number of islands. Another interpretation of the model is that there is only one island and the migrants come from the mainland where gene frequencies remain constant. The mean and variance of gene frequencies obtained for the "islands" then refer to the long term conditions on the single island.

For the island model, Wright (1931,1943) showed that for neutral alleles in a monoecious species with fully random union of gametes within subpopulations, the recurrence relation between successive values of F is

$$F^{t+1} = (1-m)^2(1 + (2N-1)F^t)/2N \qquad \ldots\ldots(2.20)$$

and at equilibrium,

$$F^{\infty} \doteq 1/(1+4Nm) \qquad \ldots\ldots(2.21)$$

where F^t is the value of F in the t-th generation.

However, inbreeding theory could be applied only to neutral genes, and so to obtain more general expressions for the amount of genetic differentiation between islands, Wright (1929,1931,1938,1969) derived frequency distributions for q (denoted f(q)) from which σ_q^2 could be calculated. These frequency distributions were based on what is now known as the diffusion equation (or Kolmogorov forward equation or Fokker-Planck equation, Kimura 1964, 1970), although Wright initially used a less elegant derivation. When an equilibrium between the gain of genes by mutation or immigration and the loss due to random sampling drift in each island is established, a stationary gene frequency distribution is reached. The frequency distribution can be obtained from the following solution to the diffusion equation:

$$f(q) = (C/V) \exp (2\int(\Delta q/V)dq) \quad . \qquad \ldots\ldots(2.22)$$

where Δq is the expected change in gene frequency in one generation, V is the expected sampling variance of the change in gene frequency (Eq. (2.1)) and C is determined by the condition.

$$\int_0^1 f(q)\, dq = 1 \qquad \ldots\ldots(2.23)$$

The probability that an island will have a gene frequency in the range q_1 to q_2 is given by $\int_{q_1}^{q_2} f(q)\, dq$.

As an example of the application of Eq.(2.22), the gene frequency distribution for the island model with immigration can be obtained from

$$\Delta q = m(\bar{q} - q)$$

$$V = q(1-q)/2N$$

Application of Eq. (2.22) and condition (2.23), yields

$$f(q) = C\, q^{4Nm\bar{q}-1}(1-q)^{4Nm(1-\bar{q})-1} \qquad \ldots\ldots(2.24)$$

where $C = (\Gamma(4Nm))/\Gamma(4nm\bar{q})\Gamma(4Nm(1-\bar{q}))$

and $E(q) = \bar{q}$

$$\sigma_q^2 = \bar{q}(1-\bar{q})/(4Nm+1)$$

From Eq. (2.19),

$$F = 1/(4Nm+1)$$

as in Eq.(2.21).

Wright (1931,1943 and 1969) applied equation (2.22) to a variety of evolutionary situations (involving mutation, immigration and selection) to obtain the ratio $\sigma_q^2/\bar{q}(1-\bar{q})$ as a measure of the amount of genetic differentiation between subpopulations of the island model. Kimura (1964, 1970) has reviewed the many applications of the diffusion equation in population genetics.

Wright (1931) referred to the fact that the value of N in Equations (2.1), (2.20), (2.21) and (2.24) might be difficult to assess in practice. The effective value of N (denoted N_e) in a subpopulation is usually less than the number that can be counted. Later, (Wright, 1938), he refined the concept of the effective size of a subpopulation or population by providing formulae for certain well-defined cases where a population deviated from the ideal random mating population of monoecious individuals. If there are two sexes with N_m males and N_f females,

$$N_e = 4N_m N_f/(N_m + N_f) \qquad \ldots\ldots(2.25)$$

so that an unequal sex ratio will reduce the effective number. If there is random mating, then the number of gametes contributed to the next generation by each parent will have a Poisson distribution with mean and variance (σ_k^2) of two. If non-random mating amongst N_o parents makes σ_k^2 different from two, then

$$N_e = (4N_o-2)/(2+\sigma_k^2) \qquad \ldots\ldots(2.26)$$

Because most natural populations are known to fluctuate in numbers, it is important to know N_e when population numbers change each generation. Wright (1938) showed that when the population size (N_i) varied over n generations, N_e was the harmonic mean of the N_i's:

$$N_e = n/(\sum_{i=1}^{n} 1/N_i) \qquad \ldots\ldots(2.27)$$

Smaller values of N_i have a greater effect on N_e than larger values, so the smallest N_i in a sequence of generations is more important than the largest in determining random sampling drift during the sequence.

Wright's initial description of the effective number in a population was found to be insufficiently precise. Kimura and Crow (1963) distinguished between inbreeding and variance effective numbers. The inbreeding effective number in a population is the number in an idealised monoecious

population with the same rate of inbreeding ($F^{t+1} - F^{t}$). The variance effective number is the number in a monoecious population with the same sampling variance for the expected change of gene frequency (c.f.Eq.(2.1)). Inbreeding effective numbers depend upon the number of parents or of grandparents (when self-fertilisation is impossible) but variance effective numbers depend upon the number of progeny.

2.2.2 The Isolation by Distance Model

Wright realised the artificiality of the island model, particularly the fact that it ignored geographic proximity between subpopulations which causes a high migration rate between nearby subpopulations and negligible migration between distant ones. The "isolation by distance" model was developed (Wright, 1943a) to cater for the other extreme type of population structure: a population with a uniform density of individuals over all parts of its range so that the only reason for the isolation of one part of the population from another is the distance between them.

In a uniformly dispersed population there are no convenient geographical or behavioural (e.g. territoriality) boundaries between subpopulations. Choosing the basic "unit" is a difficult problem. Wright's solution to this problem was based on the proposition that in a random mating monoecious subpopulation of N individuals, the probability that two gametes chosen at random will have come from the same individual is 1/N. He used this fact to derive an equation for the maximum area containing N individuals within which the subpopulation could be regarded as random mating on the grounds that an individual situated in the middle of the area had a chance of 1/N of coming from two gametes from the same parent (Wright, 1943a, 1946, 1951, 1969). This basic unit within the continuous population is now called the "neighborhood" (Wright, 1946, 1969) but was earlier called the "parental group" (Wright, 1943a) or "panmictic unit" (Dobzhansky and Wright, 1943).

When the frequency distribution of the distances between parental and offspring birthplaces and the population density are known, the neighborhood size can be calculated. For example, when the frequency distribution of parent-offspring distances is Normal, $N(o,\sigma^2)$, and population density is d, the number of individuals in a neighborhood is $3.5450\sigma d$ in a linear population and $12.5666\ \sigma^2 d$ in an area (or two dimensional) population (Wright, 1946, 1969). The linear population represents a shoreline or riverbank species whilst the area population could be a uniform stand of trees in a forest or grasses on a prairie.

Wright used the isolation by distance model to investigate the amount of genetic differentiation between randomly chosen subdivisions of a population. In so doing, he derived the basic equation for fixation indices (or F-statistics) in subdivided populations (see derivation in Chapter 3):

$$(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST}). \qquad \ldots\ldots(2.28)$$

F_{IT} is the correlation between uniting gametes in individuals relative to gametes in the total population, F_{IS} is the average correlation between uniting gametes within individuals relative to gametes in their subdivision and F_{ST} is the correlation between random gametes within subdivisions, relative to gametes of the total population (Wright, 1969). Wright did not initially recognise F_{ST} as a valid fixation index and referred to it as $\sigma^2_q/\bar{q}(1-\bar{q})$ (c.f. Equation 2.19).

Wright (1943a,1946,1969) conducted his analysis of genetic differentiation between subdivisions of a uniformly distributed population by calculating F_{ST} for subdivisions containing various numbers of neighborhoods. His major conclusions were: (a) the amount of random differentiation of neighborhoods was much greater in linear than in area populations; (b) in a linear population there can be marked random differentiation between

neighborhoods in which N is several thousand; (c) in area populations, there is virtual panmixia if the neighborhood size is greater than 1,000 and appreciable differentiation only if N<200; (d) differentiation of neighborhoods extends to differentiation of random subpopulations containing many neighborhoods.

2.2.3 The Probability of Identity by Descent.

Malecot (who wrote mainly in French, so only his 1965, 1969 works have been read; but see also Morton, 1969a, 1969b) developed the concept of the probability of identity of alleles by descent. He showed that Wright's inbreeding coefficient, F, was also the probability that the two homologous genes in a diploid individual were identical by descent from a common ancestor. He also introduced the coefficient of kinship (denoted f_{ij}; also known as the coefficient of coancestry (Malecot,1969) and coefficient of parentage (Maynard Smith, 1968)). This is the probability that homologous genes chosen randomly from individuals i and j are identical by descent from a common ancestor. The coefficient of inbreeding is naturally the average of f_{ii} over all individuals in the population.

From the definitions of F and f_{ij},

$$F^t = (1-m)^2 f_{ij}^{t-1} \qquad \ldots\ldots(2.29)$$

where f_{ij}^{t-1} is the mean coefficient of kinship in generation (t-1) and m is the proportion of genes in the population replaced each generation by unrelated genes because of mutation or immigration.

Malecot's most important contribution to the theory of subdivided populations was his analysis of the relationship between f_{ij} and the geographic distance (r) between i and j. By using advanced mathematical methods he obtained the limiting rates of decline of f_{ij} for large values of r. Denoting the expected value of f_{ij} when i and j are r units of

distance apart (r assumed large) by f(r), Malecot (1965,1969) derived the relationship between f(r) and r in a uniformly distributed population with isotropic migration. In Malecot (1965), f(r) was studied over an area population where the variance of the distance between birthplaces of mates (marital distance) is σ_m^2. When the distribution of marital distances is normal, $N(o,\sigma_m^2)$, then for large r,

$$f(r) \propto \exp((-r/\sigma_m)\sqrt{8u}) \qquad \ldots\ldots(2.30)$$

where u is the mutation rate. In a completely random mating monoecious population, f(o) is the same as the inbreeding coefficient. Malecot showed that approximately,

$$f(o) \doteq 1/(1 + 2\pi\sigma_m^2 \, d/\log (1/2u)) \qquad \ldots\ldots(2.31)$$

The same results are obtained if the probability distribution of marital distances is a "K-distribution" instead of a normal distribution. Malecot (1965) also considers the value of F when matings of high consanguinity are excluded and F no longer equals f(o). This restriction has little effect on the equilibrium value of F.

In Malecot (1969) a different genetic model is used. Instead of supposing that mutation is always giving new alleles, a mutational balance between two alleles at a locus is considered. The expected change in gene frequency due to recurrent mutation is given by Equation (2.4). From earlier discussions, we know that this mathematical model will also apply when there is immigration from another population or heterosis (but only approximately). Another difference from the previous model is that migration is now measured by σ_i^2, the variance of the distance of migration in each generation along each geographical axis.

For an area population, letting k=u+v,

$$f(o) \doteq 1/(1-8\pi\sigma_i^2 \, d/\log(2k-k^2)) \qquad \ldots\ldots(2.32)$$

and

$$f(r) \propto (1/\sqrt{r}) \exp(-(r/\sigma_i)\sqrt{2k}) \qquad \ldots\ldots(2.33)$$

For a linear population, Malecot (1969) showed that the same formulae for f(o) and f(r) apply whether the population is uniformly distributed or occupies a set of discrete colonies (i.e. distributed discontinuously). These formulae are:

$$f(o) \doteqdot 1/\ (1 + 4N\sigma_i\sqrt{2k}) \qquad \ldots\ldots(2.34)$$

and

$$f(r) \doteqdot f(o)\ \exp((-r/\sigma_i)\sqrt{2k}) \qquad \ldots\ldots(2.35)$$

The value of N in Equation (2.34) appears to be similar to Wright's neighborhood, but this is not made clear. Malecot (1969) comments that f(r) for linear and area populations depends upon the ratio r/σ_i (a standardised distance) and k, but not upon the population density, d.

All the equations for f(r) suffer from the limitation that they are derived for large r and do not predict what happens when r is small. Imaizumi *et al.* (1970) used computer simulations to examine the decline in f(r) for small distances in populations of discrete colonies. They found that

$$f(r) \propto \exp(-br) \qquad \ldots\ldots(2.36)$$

where b is a constant, was a good approximation for all populations, regardless of dimensionality, when the magnitude of f(r) was not negligible. Equation (2.33) is probably not valid for small values of r because it contains a $1/\sqrt{r}$ term whilst Equations (2.30) and (2.35) are satisfactory.

2.2.4 The Stepping Stone Model

Kimura (1953) suggested a model of population structure called the stepping-stone model. This restricts the population to discrete colonies (arranged in a line or on a lattice) with migration only between colonies that are immediate neighbors. This model was also used by Crosby (1960) in computer simulation. The mathematical consequences of the model when there are an infinite number of colonies were developed by Kimura and Weiss (1964), who obtained equations for the correlation in gene frequencies

(denoted c(r)) between colonies r steps apart in one, two and three dimensional populations. When the proportion of immigrants from neighboring colonies is m (isotropic migration assumed), N is the variance effective size of each colony, and k is (u+v) or "long range migration" (Kimura and Weiss used m_∞ as their symbol), then in a linear population,

$$\sigma_q^2 = \bar{q}(1-\bar{q})/(1 + 4N\sqrt{m2k}) \qquad \ldots\ldots(2.37)$$

and $$C(r) \propto \exp(-r\sqrt{2k/m}) \qquad \ldots\ldots(2.38)$$

provided m>>k.

Noting that $\sigma_i^2=m$ and $f(o) = \sigma_q^2/\bar{q}(1-\bar{q})$, then we can see that Equations (2.34) and (2.37) are identical, even though they were derived by quite different methods. Because f(r) can also be interpreted as a correlation, Equations (2.35) and (2.38) are essentially the same.

For an area population,

$$C(r) \propto (1/\sqrt{r}) \exp(-r\sqrt{4k/m}) \qquad \ldots\ldots(2.39)$$

so that Equations (2.33) and (2.39) are identical since $\sigma_i^2 = m/2$. Again, however, it should be emphasised that these equations were derived for r large, and the work of Imaizumi *et al.* (1970) suggests that the $(1/\sqrt{r})$ term in Equation (2.39) may not be accurate when r is small.

Bodmer and Cavalli-Sforza (1968) and Maruyama (1969,1970d,1970e,1971) considered the stepping-stone model when there are a finite number of colonies. Provided that the number of colonies is not unrealistically small, f(o) and C(r) are not greatly changed.

A difference between the methods of Wright (based on path coefficients) and those of Malecot and Kimura should be noted: the former assumes neutral alleles, with or without mutation, but the latter must include mutation or long range migration if their equations are to have non-trivial solutions. The reason for this difference does not appear to have been explained.

2.2.5 The Migration Matrix Model

According to Morton (1969b), Malecot was the first to use a migration matrix to deal with a subdivided population in a wholly general way. Consider a population with n subpopulations and let p_{ij} be the probability that an individual reproducing in the i^{th} subpopulation came from subpopulation j. The matrix of p_{ij} values (denoted $p = p_{ij}$) is called the migration matrix. This is used to find f_{ij} (in this case interpreted as the probability of drawing at random homologous genes identical by descent from subpopulations i and j) in any generation after the population is established. The island and stepping-stone models of population structure are special cases of the migration matrix where great algebraic simplification is possible.

Bodmer and Cavalli-Sforza (1968) used the migration matrix to find the variance in gene frequencies between subpopulations and the correlation in gene frequencies between particular subpopulations. To overcome the problem that the sampling variance of a gene frequency is a function of the gene frequency (Eq. (2.1)), they used the angular transformation Θ = arc sine$\sqrt{q}$. The sampling variance of Θ is approximately independent of Θ unless the gene frequency is close to 0 or 1 (Fisher and Ford,1947).

The migration matrix method is potentially very powerful because any pattern of migration between subpopulations and any variation in the size of subpopulations can be handled in an analysis of genetic differentiation due to random sampling drift between subpopulations. But, before it can be applied to a real population, an enormous amount of work will be required to estimate p_{ij} and N_e for each subpopulation to sufficient accuracy.

2.3 EVOLUTION IN A SUBDIVIDED POPULATION

2.3.1 Neutral Genes

The effect of population subdivision on the dynamics of neutral genes

was first dealt with by Wright (1943a) when he compared the variance effective number in a population that has fixation index F due to consanguineous matings (but no population subdivision), with that in a subdivided population with inbreeding due to population subdivision (F_{ST}). In the inbred, but not subdivided, population of N_T diploids the variance in the expected change in gene frequencies is

$$E(\delta q)^2 = q(1-q)(1+F)/2N_T \quad \ldots\ldots(2.40)$$

whereas in the subdivided population,

$$E(\delta q)^2 = q(1-q)(1-F_{ST})/2N_T \quad \ldots\ldots(2.41)$$

Subdivision reduces the sampling variance, increasing the variance effective number of the population and lowering the rate of loss of neutral alleles from the population. Maruyama (1972) approached this problem from quite a different angle. When f_o is the average probability of identity of allelic state (or of constitution) within each subdivision and $\bar{f}$ is the mean probability of identity of allelic state of two randomly chosen alleles from the whole population,

$$E(\delta q)^2 = q.(1-q)(1-f_o)/2N_T(1-\bar{f}) \quad \ldots\ldots(2.42)$$

The relationship between Equations (2.41) and (2.42) will be considered further in Chapter 5.

The effective population size for the island model with a finite number (n) of subpopulations, each of N haploids, was studied by Moran (1959,1962) and Pollak (1968). The result of Pollak is more general. He let the probability that a gene is exchanged between any pair of subpopulations be

$$\mu = C/N$$

Hence the frequency of immigration (from all other n-1 subpopulations) is

$$m = (n-1)\mu$$
$$= (n-1)\ C/N \quad \ldots\ldots(2.43)$$

The actual number in the population is nN and Pollak denoted the effective number by nN/λ, where approximately,

$$\lambda = n^2c + n/2 - (n/2)\sqrt{1 + 4n^2c^2 + n(n-2)c} \qquad \ldots\ldots(2.44)$$

Comparison of Pollak's work with Equation (2.41) suggests that λ should be very similar to $1-F_{ST(n)}$, where $F_{ST(n)}$ is the equilibrium value for an island population of n islands instead of an infinite number of islands (F_{ST}). As a first approximation, $F_{ST(n)}$ will be about $(n-1)/n$ times the value of F_{ST}, where

$$F_{ST} = (1-m)^2/(N-(N-1)(1-m)^2) \qquad \ldots\ldots(2.45)$$

Substituting for m (Equation 2.43) and ignoring terms of order $(1/N^2)$, a little rearrangement reveals

$$\begin{aligned} 1-F_{ST(n)} &\doteq (2(n-1)C + 1/n)/(2(n-1)C + 1) \\ &\doteq (2Nm + 1/n)/(2Nm + 1) \qquad \ldots\ldots(2.46) \end{aligned}$$

Numerical calculations indicate that λ and $1-F_{ST(n)}$ are similar (e.g.n=9, C=.125 : λ=.7095; $1-F_{ST(n)}$=.704). Thus the results of Wright (1943a) and Pollak (1968) are in agreement.

The expected number of migrants entering each subpopulation is (n-1)C, which is large if n is large unless C is very small. Moran (1959,1962) and Ewens (1969) restrict this model to cases where $C \geq 1$, which implies many immigrants into each subpopulation and can even lead to the situation where most individuals in each subpopulation are immigrants. This restricted model is probably a very poor representation of reality because in most natural populations the number of subpopulations is large but the number of immigrants per generation is small. Hence the comments that Moran (1959, 1962) and Ewens (1969) make about the effects of population subdivision on evolution can probably be ignored.

Pollak (1968) allows for the more realistic case where $0 \leq C \leq N/(n-1)$ and he concludes that if $C \ll 1$, then subdivision will markedly slow down

the rate of approach to homozygosity in the population (in the absence of mutation).

Maruyama (1970a,1970c) dealt with the rate of loss of heterozygosity (λ) in linear and circular stepping-stone models of finite length (n colonies, each of N diploids).

For the linear population,

when $Nm < n/5$; $\lambda \doteqdot 5m/n^2$, ...(2.47a)

when $Nm > n/5$; $\lambda \doteqdot 1/2Nn$ .....(2.47b)

and for the circular population,

when $Nm < n/5$; $\lambda \doteqdot 10m/n^2$ .....(2.48a)

when $Nm > n/5$ $\lambda \doteqdot 1/2Nn$ .....(2.48b)

Since $\lambda = 1/2Nn$ in a panmictic population of Nn diploids, these results give the conditions under which these stepping-stone models behave differently from a panmictic population. Maruyama drew two conclusions from his analysis: firstly, when the number of immigrants entering each subpopulation (Nm) is sufficiently small, random sampling drift in the whole population becomes independent of the size of each subpopulation; secondly, the longer a linear population, the more likely it is to deviate from panmixis.

Maruyama (1971) analysed the two-dimensional stepping-stone model of finite size and found that the critical region in which a deviation from panmixis ($\lambda = 1/2Nn$) occurred, was $Nm < 2$. In this region, $\lambda \doteqdot m/2n$. Maruyama evaluated λ exactly for a range of values of Nm and showed that as Nm passed through the critical value (Nm=2), the change-over from panmictic to subdivided ($\lambda \doteqdot m/2n$) conditions took place quickly.

Another way to examine the significance of population subdivision is to calculate the effective number of alleles (n_e) (Kimura and Crow,1964) which is defined as $1/\bar{f}$ (in this case, $\bar{f}$ is the probability of identity by descent of two alleles randomly chosen from the whole population and the "infinite alleles" model of mutational evolution in which new alleles

are produced at a rate u, is assumed). Maruyama (1970f) found for the island model

$$n_e = 1 + 4Nnu + nu/m \quad \ldots\ldots(2.49)$$

although this formula may be slightly in error because a term was omitted in part of its derivation (Latter 1973a; Cockerham, 1973; compare also with Morton, 1969b). For the circular stepping-stone model

$$n_e = 4Nnu + n\sqrt{u/2m} \quad \ldots\ldots(2.50)$$

and for linear and area stepping-stone models,

$$n_e = 4Nnu/(1-f_o) \quad \ldots\ldots(2.51)$$

where f_o is evaluated from rather complicated equations. In a panmictic population, $n_e = 1 + 4Nnu$ so Equations (2.49) and (2.50) suggest that $m < 1/2N$ is required before n_e rises much above the level in a panmictic population.

Another feature of subdivided populations was considered by Kimura and Maruyama (1971) and Maruyama (1971): the geographical pattern of a neutral polymorphism. Based on an analysis of the area stepping-stone model, they assert that when $Nm>1$ (and especially if $Nm>4$) there will be similar gene frequencies in all subpopulations, but if $Nm<1$, very different gene frequencies will occur in well separated parts of the population. Regions of high and low gene frequencies will be connected by clines in gene frequency. Hence observations on gene frequencies in natural populations that show either uniform gene frequencies over wide areas, or rapid changes in gene frequency across a landscape, cannot be taken as evidence for or against selection. These differing geographic patterns of gene frequencies could be due either to large or small values of Nm, or to selection that is homogeneous over a wide region or differs between subpopulations. Observations of gene frequencies in subpopulations of a natural population will not help to discriminate between the "neutralist"

or "selectionist" views on the cause of most polymorphisms (Kimura and Maruyama, 1971).

To summarise, the effect of population subdivision is to slow the random sampling drift of a neutral allele, but for this effect to become significant the amount of gene flow between subdivisions of the population must be small (the number of immigrants per generation should be about one or fewer). A linear population is more likely to deviate from panmixis than an area population.

2.3.2 Selection

The role of selection in producing genetic differentiation of a subdivided population has been considered in three main ways: localised variation in gene frequencies due to selection pressures that change from place to place; clines formed between two regions where different alleles are favoured; and the wave of advance of a newly arisen, selectively advantageous gene. The stability of a polymorphism in a population inhabiting diverse environments has also been a popular topic.

2.3.2.1 Selective Differentiation of Subpopulations

Haldane (1930) and Wright (1931) studied the interaction between selection and immigration in an isolated subpopulation receiving migrants from a larger source population. When selection and immigration have opposite effects on the gene frequency, an equilibrium may be established, but the exact equation for this equilibrium depends upon the type of selection (e.g. selection on a recessive, dominant or partially dominant gene). An obvious generalisation is that the gene frequency difference between the source population and the recipient subpopulation is a function of the relative magnitudes of m and the selective coefficient (s). If $m > s$, then immigration overwhelms selection and there is only slight differentiation of the subpopulation, but if $s > m$, then migration may only

serve to prevent the selectively favoured allele becoming fixed in the subpopulation. When the effective number in the subpopulation is small, a deterministic analysis of the gene frequency is inadequate and a stochastic analysis is required. The frequency distribution of gene frequencies in the subpopulation can be obtained from the solution to the diffusion equation (Equation (2.22)).

Wright (1943a) attempted an analysis of differentiation of subdivisions by selection and random sampling drift. He gave three models to show that various patterns of differentiation could arise from different circumstances.

(a) Islands with differing population sizes but the same number of immigrants each generation. Nm is identical for all islands so that differentiation due to random sampling drift (given approximately by $1/(4Nm+1)$, Eq.2.21) is the same in all islands. But for the same intensity of selection (s), differentiation will be greater on larger islands because the m:s ratio is smaller. Hence genes that are subject to different selection pressures on different islands will have a greater value of F_{ST} over large islands than over small islands.

(b) The size of each subpopulation is proportional to area and the number of immigrants is proportional to the length of the boundary around each subpopulation (i.e. $Nm \propto \sqrt{N}$). Here there is more random sampling drift in smaller subpopulations than in larger ones, and again the possibilities for selective differentiation are greater in larger subpopulations.

(c) If both subpopulation size and immigration rate are proportional to the area (i.e. m is constant), there is no relationship between subpopulation size and selective differentiation, but smaller subpopulations will be subject to more random sampling drift and hence "non-adaptive" differentiation than larger ones (Wright, 1943a).

Cain and Sheppard (1954), in an attack on the theory that a greater

variance in gene frequencies (and hence, F_{ST}) between small colonies of the snail *Cepaea* than between larger colonies was evidence for neutrality of the genes concerned, proposed another model of geographic selection. They noted that the environment in which the snails lived was very heterogeneous (made up of diverse patches, in the sense of Levins, 1968). Larger colonies extend over more patches of differing environmental conditions than small colonies. This observation implied that larger colonies would be more alike in average environmental conditions than small colonies and hence small colonies would be subject to a greater diversity of selection pressures than larger ones. Under this model, both adaptive and non-adaptive differentiation will be greater in smaller colonies.

These four different models for the interaction of selection and random drift in subdivided populations and the differing predictions obtained from each, show how difficult it is to interpret gene frequency data from natural populations as evidence for or against selection. Lamotte (1959) in a reply to Cain and Sheppard (1954) clearly identifies the problems of discerning the roles of selection and of random sampling drift in promoting differentiation of colonies of snails: "... the environmental diversity provokes a genetic diversity in the composition of the populations. As shown by experimental studies micro-climatic conditions are doubtless effective factors determining the genetic equilibrium. Selection by predators is also effective, at least in some cases. However, when one attempts to assess the cumulative action of these factors in the course of time, there always remains some unexplained residual diversity."

2.3.2.2 Selective Clines

The first analysis of a selective cline was by Haldane (1948) for a locus with a dominant and a recessive allele. The habitat is assumed to change only at a sharp boundary between two contrasting

environments which are otherwise uniform. On one side of the boundary, the recessive phenotype has a fitness (1+K) relative to the dominant phenotype, but on the other side it has relative fitness (1-k). The gene frequency at the boundary was found to depend upon the ratio, K/k. From estimates of the variance of the dispersal distance each generation (σ_d^2) and the interquartile distance of gene frequencies (with distance scaled in units of σ_d), the selection differential across the cline can be estimated. Haldane (1948) applied his model to a cline in the *Wc* gene in *Peromyscus polionotus*, calculating that K and k need only be about 0.001 to explain the observed change in gene frequencies.

Fisher (1950) introduced a different model of a cline: he assumed two codominant alleles and an environment that changed geographically such that the advantage (or disadvantage) of an allele at a point was a linear function of the distance between the point and the boundary at which both alleles were neutral. A method of determining the ratio between the rate of diffusion of genes through the population and the gradient in selective values was provided.

Clarke (1966) examined the effects of modifier genes on the slope and position of a cline. His results show that modifier genes can steepen a cline and even put a step in a previously smooth cline.

Hanson (1966) derived a more general formula for a cline and applied it to the case of a gene favoured in a selective "pocket" of a uniformly distributed population, but disadvantageous elsewhere. He showed that the size of the pocket and the selective advantage had to exceed certain bounds before the gene could be maintained in the pocket against the influx of other alleles from surrounding areas. This problem was also touched upon by Skellam (1951).

Jain and Bradshaw (1966) made computer simulations of different models of clines, showing that marked gene frequency changes could occur

over short distances, even in the face of heavy gene flow, with only moderate selection pressures.

All these mathematical analyses of theoretical clines assume that that environment changes in a remarkably precise fashion, that there is a simple population structure (usually a uniform population density) and that selection acts in a very simple manner. Such idealised populations are unlikely to be met with in nature and this fact makes it difficult to use these models to analyse what is going on in a given situation, although of course, these models do show what could happen, given the appropriate circumstances. The only attempt to apply any of these models to real data is that of Kettlewell and Berry (1961) who analysed the melanic cline in the moth *Amathes glareosa* in Shetland, using Haldane's (1948) model. Their calculations show that selective differences of less than 1% could create the cline.

Bishop (1972) attempted to estimate selective values and dispersal rates throughout a melanic cline in the moth *Biston betularia* near Liverpool. When these estimates were used in a computer simulation of the cline, the simulated cline differed from the real cline, for reasons which could not be discerned. Evidently, more factors were involved in the maintenance of the cline than Bishop was able to measure.

2.3.2.3 Stability of polymorphisms due to geographical selection

The above discussion of the balance established between opposing selection pressures in different subpopulations suggests that if the population is treated as a whole, a stable polymorphism has been established (some subpopulations may be monomorphic, however). Levene (1953) studied the stability of a polymorphism established by opposing selection pressures, using a simple model in which mating occurs at random in the population but selection takes place in different sub-

populations (called "niches" by Levene). Provided that the mean fitness of the heterozygote over all subpopulations was greater than the mean fitness of either homozygote, there should be a stable equilibrium. Levene's model has been revised, modified and improved many times (Parsons, 1963; Maynard Smith, 1966,1970; Deakin, 1966,1968,1972; Prout,1968; Bulmer, 1972; Karlin and McGregor, 1972; Gillespie, 1974; Strobeck,1974; Christiansen, 1974) but without altering the basic conclusion that geographical selection in opposition in different subpopulations can maintain a polymorphism.

Moran (1962) discussed a simple population with two subpopulations and concluded "So long as selection operates in opposite directions in the two subpopulations a stable equilibrium is possible". But Karlin and McGregor (1972), in a more rigorous analysis of the two subpopulations model, found that this intuitive remark was not always correct. In particular, if a dominant allele is favoured in one subpopulation and selected against in another, the migration rate is great enough and the selection coefficients fall within certain limits, then no stable polymorphism may be possible. This result reinforces Hanson's (1966) discovery that the area of a selective pocket has to exceed a minimum size before a locally favoured allele can be maintained.

When gene flow is not excessive and selection coefficients do not fall within certain limits, it appears that a subdivided population has a greater chance of establishing a stable polymorphism than a panmictic population.

2.3.2.4 The Wave of Advance of an Advantageous Gene.

Yet another aspect of selection in subdivided populations was dealt with by Fisher (1937) when he discussed the wave of advance through a population of an advantageous gene. This paper takes cognizance of the fact that in a subdivided population (Fisher considers isolation by

distance), a newly formed advantageous gene will not increase in frequency in all parts of the population simultaneously, but will become fixed in the area where it first arose and then spread through the population as a wave. Fisher considered a codominant gene with selective advantage, s, in a uniform linear population, with the variance of the distance between parental and offspring birthplaces given by σ_d^2. The velocity of any point on the wave is

$$V = \sigma_d\sqrt{2s}$$

and the length of the wave is proportional to $\sigma_d/\sqrt{2s}$. For example, if $s = .01$ and $\sigma_d = 100$ yards, then $V = 14$ yards per generation. Clearly, for species with limited dispersal abilities, a slightly advantageous gene could take hundreds or thousands of generations to spread through the whole population. Fisher (1937) makes the point "... the number of such waves of selective advance, simultaneously in progress, must be large."

Also, the rate of change of gene frequency at any spot on the wave is very slow and may not be detectable until many generations have passed. Thus, if observations on a population, lasting only a few generations, reveal a cline in gene frequencies, it will not be possible to distinguish between the three possibilities of (a) a neutral gene that just happens to have a cline, (b) a cline maintained by opposing selection pressures across an environmental gradient, and (c) a wave of advance of an advantageous gene.

Skellam (1951) obtained similar results with more complex models.

CHAPTER 3

HETEROZYGOTE FREQUENCIES IN SMALL SUBPOPULATIONS*

3.1 INTRODUCTION

Levene (1949) noted that when a finite sample of diploids is drawn from a population in Hardy-Weinberg equilibrium, then on average, the frequency of heterozygotes in the sample will exceed the Hardy-Weinberg expectation for the sample. This was confirmed by Cannings and Edwards (1969) using a different proof. Haldane (1954) and Smith (1970) provided statistics to test the genotype frequencies in a finite sample for deviations from Hardy-Weinberg frequencies in the population from which the sample was drawn, taking account of the effect of sampling on genotype frequencies. However, Robertson (1965) showed that it was not valid to expect Hardy-Weinberg frequencies in a population if there were a small number of parents for the individuals in the population. Robertson demonstrated that if the variance effective number of parents was N, then because of the gene frequency difference between the sexes that arises by chance, the frequency of heterozygotes in the progeny population will exceed Hardy-Weinberg expectations by a proportion $\frac{1}{2N}$, on average.

In small populations, the frequency of heterozygotes will be affected by both of the processes described, for in each generation only a finite number of progeny will be produced and only a small number of these will mate to produce the next generation. For this reason genotype frequencies in small populations (determined either by sampling or by complete census) may show an excess of heterozygotes when tested by the methods of Levene (1949), Haldane (1954) and Smith (1970) even when there are no selective differences between the genotypes. The hypothesis that we might expect to find heterozygote excesses in small populations is supported by the data of Selander (1970) and Kirby (1973) on small mouse

---

* Submitted to Theoretical Population Biology, December, 1973.

populations and a study of Yanomama Indians by Neel and Ward (1972). In all three cases there was a greater excess of heterozygotes than could be explained by Levene's (1949) correction for sample sizes.

Wright (1965, 1969) has pointed out that in many circumstances the fixation index within subpopulations should be negative, indicating a heterozygote excess. This paper uses Wright's F-statistic terminology (but a slightly different approach) to obtain quantitative estimates of the frequency of heterozygotes in small populations. It will be shown that there are two definitions of the mean fixation index within subpopulations and that one of these definitions is more useful than the other.

The emphasis will be on a population with separate sexes because this is the situation most commonly met with in practice though often neglected in the literature.

3.2 F-STATISTICS FOR SUBDIVIDED POPULATIONS

The genetic structure of a subdivided population can be described by five F-statistics, which are a set of correlations and covariances (Wright, 1951, 1965; Jain and Workman, 1967; Barrai, 1971). The three basic F-statistics (F_{IT}, F_{ST}, F_{IS}) have been known for some time (Wright, 1951, 1969) but an additional two (F_{GM}, F_{NM}) were recently introduced by Barrai (1971). It will be shown that these two are unnecessary.

The correlation between uniting gametes (or fixation index) (Wright, 1921, 1969) in a population (or subpopulation) is a measure of the deviation of genotypic frequencies at a given locus from the Hardy-Weinberg values.

If H_o is the observed frequency of heterozygotes for two codominant alleles and H_e is the frequency predicted by the Hardy-Weinberg law, then the fixation index, F, is $(H_e-H_o)/H_e$. If the frequency of the two alleles

is p and q, then F is related to H_o, p and q by $H_o = 2pq(1-F)$.

Wright (1951, 1969) described the method for applying F-statistics to subdivided (or hierarchical) populations. The F-statistic for the total population, denoted F_{IT}, is the fixation index for the total population so that $H_o = 2\bar{p}\bar{q}(1-F_{IT})$, where $\bar{p}$ and $\bar{q}$ are weighted mean allele frequencies and H_o is the frequency of heterozygotes in the whole population. The total population may be a real aggregation of subpopulations or may be an imaginary ensemble of replicate subpopulations. Both concepts are useful.

For a particular subpopulation (say, the i^{th} one) the F-statistic is F_{IS_i}, which is the fixation index for that subpopulation. The average F-statistic over all subpopulations (F_{IS}) is obtained by weighting the F_{IS_i} value for each subpopulation, usually by the size of the subpopulation (although other weighting systems may be appropriate for different circumstances). In the literature there exist two different definitions of F_{IS} which will be denoted by $\bar{F}_{IS}$ and $\hat{F}_{IS}$.

A third F-statistic, F_{ST}, arises because of the variation in gene frequencies between subpopulations due to the Wahlund effect. In general, $F_{ST} = \frac{-\sigma pq}{\bar{p}\bar{q}}$ where σpq is the covariance between the frequencies of the two alleles over all subpopulations. However, in the diallelic case ($p + q = 1$)(to which we will restrict our attention) $F_{ST} = \sigma_q^2/\bar{p}\bar{q}$, where σ_q^2 is the variance in gene frequencies over all subpopulations.

Nei (1965) provided a complete description of genotype frequencies in a subdivided population and independently, Barrai (1971) introduced two F-statistics to deal with the covariances of the subpopulation fixation index (F_{IS_i}) with the gene frequency (σ_{Fq}) and with the square of the gene frequency (σ_{Fq^2}). These F-statistics are $F_{GM} = \sigma_{Fq}/\bar{p}\bar{q}$ and $F_{NM} = \sigma_{Fq^2}/\bar{p}\bar{q}$ (Barrai, 1971).

The relationship between these five F-statistics can be obtained

by considering the two equations for the frequency of heterozygotes in the total population (H_T):

$$H_T = 2\bar{p}\bar{q}(1-F_{IT}) \quad \ldots\ldots(3.1)$$

and

$$H_T = 2\Sigma w_i p_i q_i (1-F_{IS_i}) \quad \ldots\ldots(3.2)$$

where w_i is the proportion of the total found in the i^{th} subpopulation.

The first definition of F_{IS} that will be used is the simple one of

$$\bar{F}_{IS} = \Sigma w_i F_{IS_i} \quad \ldots\ldots(3.3)$$

By noting that

$$\bar{q} = \Sigma w_i q_i \quad \ldots\ldots(3.4)$$

$$\sigma_q^2 = \Sigma w_i q_i^2 - \bar{q}^2 \quad \ldots\ldots(3.5)$$

$$\sigma_{Fq} = \Sigma w_i q_i F_{IS_i} - \bar{q}\bar{F}_{IS} \quad \ldots\ldots(3.6)$$

$$\sigma_{Fq^2} = \Sigma w_i q_i^2 F_{IS_i} - (\sigma_q^2 + \bar{q}^2)\bar{F}_{IS} \quad \ldots\ldots(3.7)$$

and equating the R.H.S. of Eq. (3.1) and Eq. (3.2), it is straightforward to obtain (Barrai, 1971)

$$1-F_{IT} = (1-F_{ST})(1-\bar{F}_{IS}) + F_{NM} - F_{GM} \quad \ldots\ldots(3.8)$$

Hence

$$\bar{F}_{IS} = \frac{F_{IT} - F_{ST} + F_{NM} - F_{GM}}{1 - F_{ST}} \quad \ldots\ldots(3.9)$$

and if $F_{NM} - F_{GM} = 0$, then

$$\bar{F}_{IS} = \frac{F_{IT} - F_{ST}}{1 - F_{ST}} \quad \ldots\ldots(3.10)$$

Eq. (3.10) is often presented with unproven assumption that $F_{NM} - F_{GM} = 0$ for subpopulations undergoing random drift. Whilst σ_{Fq} will be zero if $\bar{q} = 0.5$ and the gene frequency distribution among the different subpopulations is symmetrical about 0.5, it will not necessarily be zero when

$\bar{q}$ deviates from 0.5 and the gene frequency distribution is no longer expected to be symmetrical. Also, when $\sigma_{Fq} = 0$, it is not possible to assert that $\sigma_{Fq^2} = 0$. Thus, it will not be valid to assume that, in general, $F_{NM} - F_{GM} = 0$.

However, by redefining F_{IS}, it is possible to obtain Eq. (3.10) exactly. (I am indebted to Professor S. Wright for pointing this out.)

Let

$$\hat{F}_{IS} = (\Sigma w_i q_i p_i F_{IS_i})/(\Sigma w_i q_i p_i) \qquad \ldots\ldots(3.11)$$

and then by equating the R.H.S. of Eq.(1) and Eq.(2),

$$\hat{F}_{IS} = \frac{F_{IT} - F_{ST}}{1 - F_{ST}} \qquad \ldots\ldots(3.12)$$

Wright (1965, 1969) implied this definition of F_{IS} when he discussed the mating-type matrix model for inbreeding in sib-mating populations, but did not present it explicitly. Equations for F_{IT} and F_{ST} can be obtained for many population structures, but equations for F_{NM} and F_{GM} are not available at present. This fact makes $\hat{F}_{IS}$ a more useful parameter than $\bar{F}_{IS}$ in discussions of the expected frequency of heterozygotes within sub-populations. Some other comparisons of $\bar{F}_{IS}$ and $\hat{F}_{IS}$ are interesting and will be presented.

3.3 GENOTYPE FREQUENCIES AT DIFFERENT STAGES OF THE LIFECYCLE

Before setting out equations for F_{IT}, F_{ST} and $\hat{F}_{IS}$, it will help to clarify the approach used in this paper if it is first shown how genotype frequencies can change at different stages of the lifecycle. Consider a finite, random mating population with adults of two sexes. There are N_m males with gene frequency q_m and N_f females with gene frequency q_f. When random mating occurs, the male and female gametes unite at random to produce

a zygotic genotypic array $q_f q_m$ *AA*: $\{q_f(1-q_m) + q_m(1-q_f)\}$*Aa*: $(1-q_m)(1-q_f)$ *aa*. This genotypic array is that which would be obtained if there were an infinite number of progeny. However, this is not the genotypic array that will be observed in the N_m males and N_f females that become the adults of the next generation. These adults are a random sample from the zygotic genotypic array and will differ from the zygotic array in gene and genotype frequencies because of sampling. If migrants enter the population, they are usually envisaged as entering as adults and hence the adult phase of the lifecycle can be split into two parts: before and after immigration. Thus there may be three stages in a lifecycle at which genotype frequencies differ:

(1) breeding adults;

(2) zygotes produced by random union of gametes;

(3) adults that are a finite sample from the zygotic array.

After immigration (if it occurs) the population is back to breeding adults again.

In the theory presented in this paper, it is assumed that the population is censused for genotype frequencies at stage (3) - adults, before immigration occurs.

3.4 $\hat{F}_{IS}$ FOR AN ISOLATED DEME

An isolated deme is a random-mating subpopulation which does not receive immigrants from other subpopulations of the same species. The F-statistics will be considered over an ensemble of replicate demes with the same population parameters (e.g. size, selection pressures, mutation rates) as the particular deme study. The estimate of F_{IS} obtained in any generation is the expected value of F_{IS} in the deme under study in that generation. As is usual in inbreeding theory, the two alleles will be considered selectively neutral and the rate of mutation negligible.

The value of F_{IT} in generation t (denoted F_{IT}^{t}) can be found from the effective size of the deme in generation (t-2) (denoted N_{t-2}) and F_{IT} in previous generations with the equation

$$F_{IT}^{t} = F_{IT}^{t-1} + \frac{1}{2N_{t-2}} (1-2F_{IT}^{t-1} + F_{IT}^{t-2}) \qquad \ldots\ldots(3.13)$$

(e.g. Malecot, 1969; Wright, 1969; Crow and Kimura, 1970). If there are N_m males and N_f females in generation t, then $\frac{1}{2N_t}$ is given by $(\frac{1}{8N_m} + \frac{1}{8N_f})$.

The equation for F_{ST}^{t} can be obtained by considering the expected variance in gene frequencies (σ^2) in the ensemble of replicate demes about the initial gene frequency ($\bar{q}$) when the deme was first established. The variance in gene frequencies in generation t ($\sigma^2_{(t)}$) is

$$\sigma^2_{(t)} = E(q_i^t - \bar{q})^2 \qquad \ldots\ldots(3.14)$$

where $E(\cdot)$ indicates mathematical expectation over all demes in the ensemble and q_i^t is the gene frequency in the i^{th} deme. If only random sampling is changing gene frequencies each generation, then

$$q_i^t = q_i^{t-1} + \delta q_i \qquad \ldots\ldots(3.15)$$

where $E_d(\delta q_i) = 0 \qquad \ldots\ldots(3.16A)$

and $E_d(\delta q_i)^2 = \dfrac{q_i^{t-1}(1-q_i^{t-1})(1+F_{IS_i}^{(t-1)*})}{2N_t} \qquad \ldots\ldots(3.16B)$

(Li, 1955)

and $E_d(\cdot)$ indicates mathematical expectation within a deme. Since sampling of genotypes for the next generation occurs after mating, it is the value of F_{IS_i} in the zygotic array produced by the matings in generation (t-1) (denoted $F_{IS_i}^{(t-1)*}$) that determines the magnitude of the random drift in gene frequencies in generation t. In the zygotic array from which the individuals of generation t are picked at random, F_{ST} will remain at F_{ST}^{t-1} (because sampling in generation t has not occurred yet to increase σ^2_q)

but F_{IT} will have advanced to F_{IT}^{t} because a round of mating has occurred. F_{IT} and F_{ST} advance at different stages in each generation because mating (which advances F_{IT}) occurs before random sampling of progeny (which advances F_{ST}). Consequently

$$\hat{F}_{IS}^{(t-1)*} = \frac{F_{IT}^{t} - F_{ST}^{(t-1)}}{1 - F_{ST}^{(t-1)}} \qquad \ldots\ldots(3.17)$$

From Eq.(3.14),Eq.(3.15) and Eq.(3.16)

$$\sigma^2_{(t)} = E(q_i^{t-1} - \bar{q})^2 + E(E_d(\delta q_i)^2)$$

$$= \sigma^2_{q(t-1)} + \frac{(\bar{q}(1-\bar{q}) - \sigma^2_{(t-1)})(1+\hat{F}_{IS}^{(t-1)*})}{2N_t} \qquad \ldots\ldots(3.18)$$

and hence

$$F_{ST}^{t} = F_{ST}^{t-1} + \frac{(1-F_{ST}^{t-1})(1+\hat{F}_{IS}^{(t-1)*})}{2N_t} \qquad \ldots\ldots(3.19)$$

Substituting for $\hat{F}_{IS}^{(t-1)*}$ from Eq.(3.17) into Eq.(3.19) yields

$$F_{ST}^{t} = F_{ST}^{t-1} + \frac{1}{2N_t}(1-2F_{ST}^{t-1} + F_{IT}^{t}) \qquad \ldots\ldots(3.20)$$

This equation is notable because it contains a term in F_{IT}^{t} and hence shows that F_{ST}^{t} is not independent of F_{IT}, although from Eq.(3.13) F_{IT}^{t} is quite independent of F_{ST}.

If $F_{IT}^{o} = F_{ST}^{o} = 0$ (i.e. the population from which the founders of the deme originated was not inbred), then from Eq.(3.13) and Eq.(3.20), $F_{IT}^{(1)} = F_{IT}^{(2)} = 0$, $F_{IT}^{(3)} = F_{ST}^{(1)} = \frac{1}{2N_t}$ and subsequently $F_{IT}^{(t+2)} = F_{ST}^{(t)}$. Under these circumstances,

$$F_{ST}^{t} = F_{ST}^{t-1} + \frac{1}{2N_t}(1-2F_{ST}^{t-1} + F_{ST}^{t-2}) \qquad \ldots\ldots(3.21$$

In animal and plant breeding, Eq.(3.21) will apply when two inbred

lines are crossed and sublines are later set up, if the F_1 is taken as generation 0.

For any generation, t, $\hat{F}^t_{IS}$ can be estimated by applying Eq.(3.13) and Eq.(3.20) or Eq.(3.21) to evaluate F^t_{IT} and F^t_{ST} and substituting into Eq.(3.12). If Eq.(3.21) is applicable, then Eq.(3.12) becomes

$$\hat{F}^t_{IS} = \frac{F^t_{IT} - F^{t+2}_{IT}}{1 - F^{t+2}_{IT}}$$

and if $H^t_{IT} = 1 - F^t_{IT}$; then $\hat{F}^t_{IS} = 1 - \frac{H^t_{IT}}{H^{t+2}_{IT}}$ .....(3.22)

From Eq.(3.13) $H^t_{IT} = (1 - \frac{1}{N_{t-2}})H^{t-1}_{IT} + \frac{1}{2N_{t-2}} H^{t-2}_{IT}$ .....(3.23)

If N_t remains constant at N in all generations, then Eq.(3.23) is a homogeneous difference equation with solution $H^t_{IT} = A\lambda_1^t + B\lambda_2^t$ (e.g. Malecot, 1969), where λ_1 and λ_2 are solutions of the equation

$$\lambda^2 - (1 - \frac{1}{N})\lambda - \frac{1}{2N} = 0$$

If $\lambda_1 > \lambda_2$, then in the long term behaviour of H^t_{IT}, λ_2 can be ignored and hence $H^t_{IT} = A\lambda_1^t$.

Now, $\lambda_1 = \frac{1}{2N}(N - 1 + \sqrt{N^2 + 1})$ and hence F^t_{IS} from Eq.(3.22) tends to a limiting value of

$$\hat{F}^t_{IS} = 1 - \frac{1}{\lambda_1^2} = 1 - \frac{4N^2}{(N - 1 + \sqrt{N^2 + 1})^2}$$ .....(3.24)

as t becomes large. For the special case of a sibmating line, N=2 and $\hat{F}^t_{IS}$ tends to a limit of -0.52786.

3.5 $\hat{F}_{IS}$ IN A DEME RECEIVING IMMIGRANTS

The study of a deme receiving a steady flow of immigrants from other

subpopulations of the same species is assisted by utilising Wright's "Island" model of population structure (Wright, 1951, 1969). This model envisages an infinite number of demes receiving immigrants from a common source population with constant gene frequency $\bar{q}$, but without any migration between demes. Hence F_{IT} and F_{ST} are defined over the entire ensemble of demes relative to $\bar{q}$, and F_{IS} is the expected (or mean) value of F_{IS_i} in each deme.

The expression for F_{IT}^t is

$$F_{IT}^t = (1 - \frac{1}{N_{t-2}})(1 - m)^2 F_{IT}^{t-1} + \frac{1}{2N_{t-2}}(1 - m)^4(1 + F_{IT}^{t-2}) \quad \ldots\ldots(3.25$$

(e.g. Crow and Kimura,1970; Malecot,1969)

where m is the proportion of the gene pool in each deme contributed by immigration each generation.

The equation for F_{ST}^t must correspond in the stage of the life cycle with the equation for F_{IT}^t. The derivation of F_{IT}^t takes account of immigration in generation (t-2) and (t-1) but not generation (t), and hence Eq.(3.25) applies in generation (t) to the time after the progeny in generation (t) are obtained, but before immigration occurs in generation (t). Hence the corresponding equation for F_{ST}^t must apply after sampling but before immigration in generation (t). Let $q_{(t-1)}$ and $q^{\dagger}_{(t-1)}$ be the gene frequency in a deme before and after immigration in generation (t-1), respectively.

$$q^{\dagger}_{(t-1)} = q_{(t-1)} + m(\bar{q} - q_{(t-1)})$$

After mating and sampling of progeny in the next generation:

$$q_t = q^{\dagger}_{t-1} + \delta q^{\dagger}_{t-1}$$

where $E_d(\delta q^{\dagger}_{t-1}) = 0$

and $$E_d(\delta q^{\dagger}_{t-1})^2 = \frac{1}{2N_t} q^{\dagger}_{(t-1)}(1 - q^{\dagger}_{(t-1)})(1 + F_{IS}^{(t-1)*}) \quad \ldots\ldots(3.26$$

Now $$\sigma^2_{(t)} = (1 - m)^2\sigma^2_{(t-1)} + E(E_d(\delta q^{\dagger}_{t-1})^2))$$

$$= (1 - m)^2\sigma^2_{(t-1)} + \frac{1}{2N_t}((\bar{q}(1 - \bar{q})$$

$$-(1 - m)^2\sigma^2_{(t-1)})(1 + \hat{F}^{(t-1)*}_{IS})) \qquad \ldots\ldots(3.27)$$

$\hat{F}^{(t-1)*}_{IS}$ is evaluated after immigration and mating in generation (t-1) so that F_{IT} has become F^t_{IT}, but before sampling so that F_{ST} is $(1-m)^2F^{(t-1)}_{ST}$. Therefore,

$$\hat{F}^{(t-1)*}_{IS} = \frac{(F^t_{IT} - (1-m)^2F^{t-1}_{ST})}{1 - (1-m)^2F^{t-1}_{ST}} \qquad \ldots\ldots(3.28)$$

From Eq.(3.27) and Eq.(3.28)

$$F^t_{ST} = (1-m)^2F^{(t-1)}_{ST} + \frac{1}{2N_t}(1-2(1-m)^2F^{(t-1)}_{ST} + F^t_{IT}) \qquad \ldots\ldots(3.29)$$

$\hat{F}^t_{IS}$ can be found by substituting for F^t_{IT} and F^t_{ST} in Eq.(3.12). If the magnitudes of N and m remain constant for a sufficiently long time, the values of F_{IS}, F_{ST} and F_{IT} will approach the equilibrium values denoted $\hat{F}^{\infty}_{IS}$, F^{∞}_{ST} and F^{∞}_{IT} respectively. From Eq.(3.25) and (3.29)

$$F^{\infty}_{IT} = \frac{(1-m)^4}{2N - (2N-2)(1-m)^2 - (1-m)^4} \qquad \ldots\ldots(3.30)$$

$$F^{\infty}_{ST} = \frac{1 + F^{\infty}_{IT}}{2N - (2N-2)(1-m)^2}$$

$$= \frac{1}{2N - (2N-2)(1-m)^2 - (1-m)^4} \qquad \ldots\ldots(3.31)$$

and from Eq. (3.12)

$$\hat{F}^{\infty}_{IS} = \frac{(1-m)^4 - 1}{2N - (2N-2)(1-m)^2 - (1-m)^4 - 1} \qquad \ldots\ldots(3.32)$$

$$\doteq -\frac{(1-m)}{N} \qquad \ldots\ldots(3.32)$$

From Eq. (3.32B) it is evident that for small m, $\hat{F}^{\infty}_{IS}$ will be approximately $-\frac{1}{N}$, and that more generally, the absolute magnitude of $\hat{F}^{\infty}_{IS}$ is reduced by increasing immigration rates.

All of this theory has dealt with $\hat{F}_{IS}$ and ignored $\bar{F}_{IS}$. The value of $\bar{F}_{IS}$ cannot be estimated using the methods described above and so other methods must be tried. Two approaches will be used to compare $\bar{F}_{IS}$ and $\hat{F}_{IS}$: firstly by utilising the matrix of mating-type frequencies in the case of a sib-mating subpopulation (e.g. Fisher, 1949; Wright,1965, 1969) and secondly by computer simulation of populations of larger size.

3.6 MATRIX OF MATING TYPES FOR SIB-MATING DEMES: (a) ISOLATED DEME

In the case of an isolated deme of only one male and one female (sibs) each generation, there are six distinct mating types to consider: *AA* x *AA*, *AA* x *Aa*, *AA* x *aa*, *Aa* x *Aa*, *Aa* x *aa* and *aa* x *aa*, denoted by m_1,m_2, m_3, m_4, m_5 and m_6 respectively. If the frequency of the m_i^{th} mating type in the t^{th} generation is denoted by c_i^t, then the frequencies of the six mating types in successive generations are given in the matrix equation in Table 3.2, after substituting m=0 in the 6 x 6 matrix (e.g. Wright, 1969).

F_{IS_i} takes the values 0 in m_1 and m_6, $-\frac{1}{3}$ in m_2 and m_5, +1 in m_3 and -1 in m_4. The mean $\bar{F}_{IS}$ is obtained by weighting each mating type by its frequency,

thus $$\bar{F}^t_{IS} = c_3^t - c_4^t - \frac{1}{3}(c_2^t + c_5^t).$$

However, $\hat{F}_{IS}$ is obtained by weighting each mating type F_{IS_i} by its frequency and by $q_i(1-q_i)/(\Sigma q_i(1-q_i))$ so that

$$\hat{F}^t_{IS} = \frac{4(c_3^t - c_4^t) - (c_2^t + c_5^t)}{3(c_2^t + c_5^t) + 4(c_3^t + c_4^t)}$$

The mean gene frequency, q, is

$$q = c_1^t + .75\, c_2^t + .5(c_3^t + c_4^t) + .25\, c_5^t$$

TABLE 3.1

F-statistics for the isolated sibmating deme from the matrix model.

Generation	F_{IT}	F_{ST}	$\bar{F}_{IS}$	$\hat{F}_{IS}$
1	0.0000	0.2500	-0.2917	-0.3333
2	0.0000	0.3750	-0.4063	-0.6000
3	0.2500	0.5000	-0.2786	-0.5000
4	0.3750	0.5938	-0.2415	-0.5385
5	0.5000	0.6719	-0.1906	-0.5238
10	0.8262	0.8862	-0.0665	-0.5279
15	0.9398	0.9606	-0.0231	-0.5279
20	0.9791	0.9863	-0.0080	-0.5279
25	0.9928	0.9953	-0.0028	-0.5279
30	0.9975	0.9984	-0.0010	-0.5279

TABLE 3.2

Generation matrix equation for immigration into sibmating subpopulations ($\bar{q} = \frac{1}{2}$)

$$
\begin{vmatrix} c_1^t \\ c_2^t \\ c_3^t \\ c_4^t \\ c_5^t \\ c_6^t \end{vmatrix} = \begin{vmatrix} (1-\frac{m}{2})^4 & (\frac{1}{2}-\frac{m}{4})^2 & \frac{m^2}{4}(1-\frac{m}{2})^2 & \frac{1}{16} & (\frac{m}{4})^2 & (\frac{m}{2})^4 \\ 2m(1-\frac{m}{2})^3 & (\frac{1}{2}-\frac{m}{4}) & m(1-\frac{m}{2})(1-m+\frac{m^2}{2}) & \frac{1}{4} & \frac{m}{4} & \frac{m^3}{2}(1-\frac{m}{2}) \\ \frac{m^2}{2}(1-\frac{m}{2})^2 & \frac{m}{2}(\frac{1}{2}-\frac{m}{4}) & \frac{m^2}{2}(1-\frac{m}{2})^2 & \frac{1}{8} & \frac{m}{2}(\frac{1}{2}-\frac{m}{4}) & \frac{m^2}{2}(1-\frac{m}{2})^2 \\ m^2(1-\frac{m}{2})^2 & \frac{1}{4} & (1-m+\frac{m^2}{2})^2 & \frac{1}{4} & \frac{1}{4} & m^2(1-\frac{m}{2})^2 \\ \frac{m^3}{2}(1-\frac{m}{2}) & \frac{m}{4} & m(1-\frac{m}{2})(1-m+\frac{m^2}{2}) & \frac{1}{4} & (\frac{1}{2}-\frac{m}{4}) & 2m(1-\frac{m}{2})^3 \\ (\frac{m}{2})^4 & (\frac{m}{4})^2 & \frac{m^2}{4}(1-\frac{m}{2})^2 & \frac{1}{16} & (\frac{1}{2}-\frac{m}{4})^2 & (1-\frac{m}{2})^4 \end{vmatrix} \times \begin{vmatrix} c_1^{t-1} \\ c_2^{t-1} \\ c_3^{t-1} \\ c_4^{t-1} \\ c_5^{t-1} \\ c_6^{t-1} \end{vmatrix}
$$

and hence

$$F_{IT}^{t} = 1 - (c_2^t + c_5^t + 2c_4^t)/4q(1-q).$$

The variance in gene frequencies will be

$$\sigma^2_{(t)} = c_1^t + .5625\, c_2^t + .25\,(c_3^t + c_4^t) + .0625\, c_5^t - q^2$$

and

$$F_{ST}^{(t)} = \sigma^2_{(t)}/q(1-q)$$

$$= c_1^t + c_6^t + .25\,(c_2^t + c_5^t) \text{ when } q = \tfrac{1}{2}.$$

Some results obtained by iterating the matrix equation for (c_i^t) for 30 generations after $c_4^o = 1$ are given in Table 3.1. The values of F_{IT}, F_{ST} and F_{IS} agree with results obtained from Eq.(3.13), Eq.(3.20) and Eq.(3.24), thus providing a check on their validity. There is a marked difference between $\bar{F}_{IS}$ and $\hat{F}_{IS}$ after the first two generations. Because each term in $\hat{F}_{IS}$ is weighted by $q_i(1-q_i)$, subpopulations that are fixed for one allele or the other will not be included in this average. Hence $\hat{F}_{IS}$ is not affected by the increasing proportion of subpopulations that are homogenic (*AA* x *AA* or *aa* x *aa*) and stabilises because the relative frequencies of the heterogenic mating types stabilises. But $\bar{F}_{IS}$ is the average over all mating types (including the homogenic cases with $F_{IS_i} = 0$) and hence tends to zero as the proportion of homogenic mating types tends to unity.

3.7 SIB-MATING DEMES: (b) DEME RECEIVING IMMIGRANTS

The matrix equation for this case is given in Table 3.2. This was obtained by noting that each gamete produced within a deme at mating has a probability m of being replaced by an immigrant gamete. For example, if the gene frequency in the source population of immigrants is 0.5, then the expected genotypic ratios in the progeny of m_1 (i.e. *AA* x *AA*) will be

$(1 - \frac{1}{2}m)^2 AA : m(1 - \frac{1}{2}m)\ Aa : \frac{1}{4}m^2\ aa$, and hence the probabilities of m_1 giving rise to any particular mating type can be calculated. For simplicity, Table 3.2 was obtained assuming that the mean gene frequency of immigrants was 0.5.

This matrix equation can be solved to obtain the equilibrium values of (c_i) when m>0. For example, if m = 0.1 the equilibrium values of the (c_i) yield the following F-statistics: $F^{\infty}_{IT} = 0.3806$, $F^{\infty}_{ST} = 0.5801$, $\bar{F}^{\infty}_{IS} = -0.2268$, $\hat{F}^{\infty}_{IS} = -0.4751$, $F^{\infty}_{GM} = 0.0$ and $F^{\infty}_{NM} = 0.1043$. The values of F^{∞}_{IT}, F^{∞}_{ST} and $\hat{F}^{\infty}_{IS}$ are the same as those obtained from Eq. (3.30), Eq. (3.31) and Eq. (3.32A), and Eq. (3.32B) ($\hat{F}^{\infty}_{IS} = -0.45$) is a good approximation to Eq. (3.32A) ($\hat{F}^{\infty}_{IS} = -0.475$).

As before the values of $\bar{F}_{IS}$ and $\hat{F}_{IS}$ are different because $\bar{F}_{IS}$ is affected by the proportion of homogenic subpopulations.

The matrix of mating-types method is unwieldy with demes larger than N=2, so to compare $\bar{F}_{IS}$ and $\hat{F}_{IS}$ in larger demes it is necessary to resort to computer simulation.

3.8 COMPUTER SIMULATIONS: (a) ISOLATED DEME

A computer program was written to simulate random drift in an isolated deme. In each generation, the gene frequencies in males and females were calculated separately. Then progeny were formed by randomly selecting a gene (with replacement) from the gene pool of each sex, using pseudo-random numbers (e.g. Fraser and Burnell, 1970). Up to 1,000 replicate demes were simulated concurrently and in each generation the mean $\bar{F}_{IS}$ was calculated over all the replicate demes (with $F_{IS_i} = 0$ in demes fixed for an allele). F_{ST} and F_{IT} were calculated relative to the initial gene frequency in the demes (q = 0.5) and the initial frequency of heterozygotes (H = 0.5). Simulation runs were begun by drawing the first generation from a population in which the gene frequency was 0.5 in both sexes. Several repeat simulations with different pseudo-random

TABLE 3.3

Comparison between $\hat{F}_{IS}$ and $\bar{F}_{IS}$ in an isolated deme

Generation	N_e = 8		N_e = 16		N_e = 32	
	$\hat{F}_{IS}$	$\bar{F}_{IS} \pm$ SE	$\hat{F}_{IS}$	$\bar{F}_{IS} \pm$ SE	$\hat{F}_{IS}$	$\bar{F}_{IS} \pm$ SE
1	-0.067	-0.066 ± 0.005	-0.032	-0.033 ± 0.004	-0.016	-0.013 ± 0.003
2	-0.133	-0.128 ± 0.004	-0.064	-0.068 ± 0.003	-0.032	-0.033 ± 0.002
3	-0.128	-0.125 ± 0.004	-0.063	-0.060 ± 0.002	-0.031	-0.035 ± 0.003
4	-0.128	-0.118 ± 0.002	-0.063	-0.064 ± 0.003	-0.031	-0.030 ± 0.002
5	-0.128	-0.115 ± 0.005	-0.063	-0.061 ± 0.006	-0.031	-0.029 ± 0.002
10	-0.128	-0.093 ± 0.003	-0.063	-0.058 ± 0.004	-0.031	-0.034 ± 0.003
15	-0.128	-0.064 ± 0.003	-0.063	-0.050 ± 0.004	-0.031	-0.028 ± 0.003
20	-0.128	-0.050 ± 0.001	-0.063	-0.043 ± 0.002	-0.031	-0.032 ± 0.002
25	-0.128	-0.034 ± 0.002	-0.063	-0.034 ± 0.004	-0.031	-0.027 ± 0.001
30	-0.128	-0.030 ± 0.003	-0.063	-0.037 ± 0.001	-0.031	-0.025 ± 0.002

$\hat{F}_{IS}$ is obtained from Eq.(3.12), Eq.(3.13) and Eq.(3.21) and $\bar{F}_{IS}$ from computer simulations.

TABLE 3.4

Comparison of $\hat{F}_{IS}$ and $\bar{F}_{IS}$ in a deme with an immigration rate of 10%.

	N_e = 4		N_e = 8		N_e = 16	
Generation	$\hat{F}_{IS}$	$\bar{F}_{IS}$ SE	$\hat{F}_{IS}$	$\bar{F}_{IS}$ SE	$\hat{F}_{IS}$	$\bar{F}_{IS}$ SE
1	-0.143	-0.152 ± 0.008	-0.067	-0.071 ± 0.005	-0.032	-0.029 ± 0.003
2	-0.251	-0.235 ± 0.005	-0.120	-0.119 ± 0.003	-0.058	-0.058 ± 0.003
3	-0.236	-0.210 ± 0.009	-0.116	-0.110 ± 0.002	-0.057	-0.055 ± 0.004
4	-0.237	-0.202 ± 0.006	-0.116	-0.108 ± 0.003	-0.057	-0.061 ± 0.003
5	-0.235	-0.191 ± 0.004	-0.116	-0.107 ± 0.005	-0.057	-0.063 ± 0.004
10	-0.232	-0.178 ± 0.006	-0.115	-0.112 ± 0.008	-0.057	-0.059 ± 0.003
15	-0.232	-0.176 ± 0.004	-0.115	-0.108 ± 0.006	-0.057	-0.055 ± 0.003
20	-0.232	-0.172 ± 0.003	-0.115	-0.108 ± 0.004	-0.057	-0.056 ± 0.003
25	-0.232	-0.172 ± 0.003	-0.114	-0.112 ± 0.001	-0.057	-0.054 ± 0.003
30	-0.232	-0.180 ± 0.003	-0.114	-0.113 ± 0.004	-0.057	-0.051 ± 0.005

$\hat{F}_{IS}$ is obtained from Eq.(3.12), Eq.(3.25) and Eq.(3.29) and $\bar{F}_{IS}$ from computer simulations.

numbers were run for each set of values of N_m, N_f and number of replicate demes, to obtain means and standard errors for the estimates of $\bar{F}_{IS}$, F_{ST} and F_{IT}. The program was validated by comparing the output values of $\bar{F}^t_{IS}$, F^t_{ST} and F^t_{IT} with those predicted in the sibmating case by the matrix of mating types method.

The following deme sizes were simulated: two of each sex (N=4); four of each sex (N=8); eight of each sex (N=16) and sixteen of each sex (N=32), and some results for F^t_{IS} for N=8, 16 and 32 are given in Table 3.3. The difference between $\hat{F}_{IS}$ and $\bar{F}_{IS}$ after several generations is less marked when N is greater. $\hat{F}_{IS}$ stabilises very quickly but $\bar{F}_{IS}$ tends to zero with increasing time, although more gradually with larger N.

3.9 COMPUTER SIMULATION: (b) DEME RECEIVING IMMIGRANTS

The computer program used in the previous section was modified to allow for immigration into a deme from a source population with a gene frequency of 0.5. In each generation after the gene frequency (q) was calculated for each sex, it was changed to the expected gene frequency after immigration (q') by the equation $q' = q + m(\frac{1}{2} - q)$.

The progeny were then chosen from the modified gene pool for each sex. An immigration rate of m = 0.1 was chosen and simulations were made of demes of size N=2, 4, 8 and 16. Some of the results for $\bar{F}^t_{IS}$ and $\hat{F}^t_{IS}$ for N=4, 8 and 16 are given in Table 3.4. For this case, $\bar{F}_{IS}$ and $\hat{F}_{IS}$ are significantly different when $N_e = 4$, but they are not statistically significantly different when $N_e = 8$ or 16, although $\bar{F}_{IS}$ is clearly less than $\hat{F}_{IS}$ on average.

3.10 DEME RECEIVING IMMIGRANTS: EFFECT ON $\bar{F}_{IS}$ OF DIFFERING MEAN GENE FREQUENCIES

Any parameter used in the analysis of population structure (such as the F-statistics) which is to be of simple application, should be

TABLE 3.5

Equilibrium values of $\hat{F}_{IS}$, $\bar{F}_{IS}$, F_{GM} and F_{NM} for sibmating demes with different mean gene frequencies of immigrants and 10% immigration per generation

	Mean gene frequency of immigrants			
	0.5	0.3	0.1	0.05
$\hat{F}^{\infty}_{IS}$	-0.475	-0.475	-0.475	-0.475
$\bar{F}^{\infty}_{IS}$	-0.227	-0.191	-0.082	-0.044
F^{∞}_{GM}	0.000	-0.153	-0.307	-0.347
F^{∞}_{NM}	0.104	-0.033	-0.143	-0.166

independent of the mean gene frequency in an ensemble of subpopulations. F_{IT} and F_{ST} (and hence $\hat{F}_{IS}$) are the same for all populations with the same N, m and t and are thus satisfactory in this regard. But F_{GM} and F_{NM} will depend upon the mean gene frequency in an ensemble, and hence so will $\bar{F}_{IS}$. Toillustrate this point, the matrix of mating types for sibmating lines receiving immigrants was modified to allow calculation of $(c_{i.}^{t})$ when the immigrant's source population had any gene frequency. Table 3.5 shows the equilibrium values of $\hat{F}_{IS}$, $\bar{F}_{IS}$, F_{GM} and F_{NM} when the mean gene frequency of immigrants was 0.5, 0.3, 0.1 or 0.05. Clearly, as $\bar{q}$ deviates more from 0.5, $\bar{F}_{IS}$ tends to zero, probably because of the increasing proportion of homogenic mating types. F_{GM} and F_{NM} are also strongly affected by $\bar{q}$.

Therefore, there are two main reasons for preferring $\hat{F}_{IS}$ over $\bar{F}_{IS}$ as a parameter measuring the average fixation index within subpopulations: (1) when F_{IT} and F_{ST} can be calculated, $\hat{F}_{IS}$ is easily obtained whilst $\bar{F}_{IS}$ is not; (2) $\hat{F}_{IS}$ takes the same value for the same N, m and t regardless of $\bar{q}$ whilst $\bar{F}_{IS}$ is also affected by $\bar{q}$.

Neel and Ward (1972) unfortunately used $\bar{F}_{IS}$ in their analysis of eight loci in three South American Indian tribes by F-statistics. They neglected to calculate F_{GM} and F_{NM} and consequently did not carry out a complete analysis of their data.

3.11 THE INTERPRETATION OF ESTIMATES OF F_{IS} IN NATURAL POPULATIONS

Modern methods in gel electrophoresis have made possible extensive surveys of polymorphisms in many species. However, it is very often difficult to interpret the results of these surveys. Selander (1970) noted that in the data on two house-mouse polymorphisms (in samples collected in Texas) there was a significant regression between a coefficient related to F_{IS} and sample size (denoted N_S). The theory of F_{IS} developed in previous sections can be used to demonstrate how Selander's regression can be explained without recourse to selection and how data collected

from a number of subpopulations of various sizes can yield useful information about the population structure of a species.

If many subpopulations have been sampled and the samples differ in size, then the first test will be a check on the fit to Hardy-Weinberg frequencies. The simplest test is based on the fact that if each sample (of size N_S) were drawn from a much larger population in Hardy-Weinberg equilibrium, there should be a linear regression between the F_{IS} value and $\frac{-1}{2N_S-1}$ (obtained from $F_{IS}^{(1)}$ for an isolated deme). All that is required is to calculate the regression

$$F_{IS} = a + b\left(\frac{-1}{2N_S-1}\right)$$

and then test a for difference from zero and b for difference from one. If neither test gives a significant result, then this is good evidence that over all subpopulations sampled, there is a good fit to binomial proportions.

However some species live in local colonies that may not be very large. Each colony may be akin to the "deme with immigration" model presented above, if it has a steady trickle of immigrants from other colonies of the species. If a full census can be made of each of many colonies (that is, every individual in each colony should be typed for the locus or loci studied) then from Eq. (3.32B) we might expect there to be a linear regression between F_{IS} and $-\frac{1}{N_S}$. If only a small proportion of each colony has been censused (or immigration rates vary according to the size of each deme), this may not be true. A common problem will be that many species have overlapping generations, and the effects of this complicating factor have not been studied. Nevertheless, the regression $F_{IS} = a + b(-\frac{1}{N_S})$ could be used to test whether or not the data support the model of population structure outlined above. It should be noted that if the effective population size of each colony is less than the census size, b may be greater than unity. For example, if the effective size is $\frac{1}{3}$rd the census size, then b should be about 3 because it is the effective size that is most

important in determining F_{IS}.

Some species (especially the house mouse, *Mus musculus*) tend to form reproductively isolated subgroups (often called 'tribes' or 'families') within a colony, if the colony exceeds a certain size (Selander and Yang, 1970; Berry, 1970; Rasmussen, 1964; Anderson, 1970). Thus a sample from a medium sized or large colony may include individuals from several subgroups and it is necessary to study the changes in F_{IS} when several subgroups are sampled.

Returning to Wright's Island model, it is clear that if the variance in gene frequencies between demes is σ_q^2, then the variance in gene frequencies between demes fused in pairs will be $\frac{1}{2}\sigma_q^2$, between demes fused in triplets will be $\frac{1}{3}\sigma_q^2$, and so on. In general, if new *conglomerate* subpopulations are formed by combining a number (e.g. k) of demes such that each conglomerate consists of a proportion w_i from the i^{th} deme, then the variance in gene frequencies between such conglomerate subpopulations will be $\sum_{i=1}^{k} w_i^2\sigma^2$. If each w_i is $\frac{1}{k}$, then $\sum_{i=1}^{k} w_i^2$ is $\frac{1}{k}$. Hence in the general case where w_i is not necessarily $\frac{1}{k}$ it is convenient to define the effective number (n) of single demes in a conglomerate subpopulation by the equation $\frac{1}{n} = \sum_{i=1}^{k} w_i^2$. If the variation between single demes is described by F_{ST}, then the equivalent F-statistic for conglomerate subpopulations will be $\frac{1}{n}F_{ST}$. Since F_{IT} remains constant whether we are dealing with single demes or conglomerate subpopulations, $\hat{F}_{IS}$ in conglomerates (denoted $\hat{F}_{IS(n)}$) will be given by

$$\hat{F}_{IS(n)} = (F_{IT} - \tfrac{1}{n}F_{ST})/(1 - \tfrac{1}{n}F_{ST}).$$

If the model of a single deme with immigration is utilised, then at equilibrium F_{IT} and F_{ST} can be obtained from Eq. (3.30) and Eq. (3.31), and hence

$$\hat{F}_{IS(n)} = \frac{n(1-m)^4 - 1}{n(2N - (2N-2)(1-m)^2 - (1-m)^4) - 1} \qquad \ldots\ldots (3.33$$

For small n, $\hat{F}_{IS(n)}$ will be negative, but it will become positive as n

TABLE 3.6

$F^{\infty}_{IS(n)}$ for N=16 and three different immigration rates (m)

n	m		
	0.01	0.10	0.20
1	-0.0619	-0.0569	-0.0518
2	0.4053	0.0239	-0.0078
3	0.4814	0.0481	0.0064
4	0.5125	0.0598	0.0134
5	0.5295	0.0666	0.0175

becomes larger. $\hat{F}_{IS(n)}$ will be zero when $(1-m) = \frac{1}{n}$, showing that for larger immigration rates, n will be larger before $\hat{F}_{IS(n)}$ becomes zero. The change in $\hat{F}_{IS(n)}$ with increasing n is shown by three examples in Table 3.6. If $\hat{F}_{IS(n)}$ is plotted against n, there are three features of the regression worth noting: firstly the value of $\hat{F}_{IS(1)}$ is determined largely by the size of the single deme (Eq.(3.32B)) if m is small; secondly, the values of $\hat{F}_{IS(n)}$ for large n are determined largely by Nm since Eq.(3.30)

$$\hat{F}_{IS(\infty)} = F_{IT}^{\infty} = \frac{1 - 4m}{1 + 4Nm} ; \qquad \ldots\ldots (3.34)$$

and thirdly, the slope of the regression is inversely proportional to the immigration rate because a lower migration rate gives a greater change in $\hat{F}_{IS(n)}$ from n=1 to n=5 (see Table 3.6).

When a sample is collected by censusing a colony which contains several subgroups, this sample will be analogous to a conglomerate sub-population. It will differ slightly from the type of conglomerate analysed here because in the real situation there will be two forms of immigration to consider: immigration from other conglomerates and inter-migration between the demes within the conglomerate. The more elaborate model will not be analysed here because it will require a full analysis of two levels of population subdivision (between conglomerates and within conglomerates). As a first approximation, the simple model of the population structure of conglomerates outlined above, will be useful. The value of m in this case is poorly defined for it will have to include both immigration from other unrelated conglomerates and also the immigration into each deme within a conglomerate from other such demes. Despite these restrictions, the three main predictions of the simple model should be verifiable in natural populations.

A very real practical problem arises in obtaining the regression of $\hat{F}_{IS(n)}$ on n: that of estimating the magnitude of n in a sample. Special

care will need to be taken to estimate n before much can be claimed about the slope of the regression of $\hat{F}_{IS(n)}$ on n. In the only data available which are suitable for testing the three main predictions of the model (Selander *et al.*,1969), there is no way of estimating the value of n in each sample. Hence only the first two predictions can be tested. Selander (1970) (see also Selander and Yang, 1970) has already provided a linear regression of D on N_S using these data. As defined by Selander, D is approximately equal to $-F_{IS}$. In calculating the expected frequency of heterozygotes (h_e), Levene's (1949) correction for small samples was applied by Selander, but this is unnecessary when F_{IS} has been obtained by censusing all (or nearly all) of the individuals in a colony. If sample size (N_S) is approximately proportional to n, then from Eq.(3.33) and Table 3.6 it is clear that the regression of F_{IS} on N_S will be curvilinear and a linear regression may not provide a good fit to the data. Although these data are not suitable for calculating a regression of $\hat{F}_{IS(n)}$ on n, some of the data are potentially useful. The values of F_{IS} in the smallest sample sizes may provide an estimate of $\hat{F}_{IS(1)}$ and the values in the largest sizes may approach $\hat{F}_{IS(\infty)}$ in magnitude. If the data on the *Es-3* locus for the samples with only two alleles ($Es\text{-}3^b$ and $Es\text{-}3^c$) are taken, it is found that for $10<N_s<14$, $\hat{F}_{IS} = -0.1274 \pm 0.058$ (21 samples) and for $N_s>100$, $\hat{F}_{IS} = 0.0524 \pm 0.024$ (17 samples). Both estimates of $\hat{F}_{IS}$ are significantly different from zero, and each differs in the direction predicted by Eq.(3.33). Unfortunately, data on the *Hbb* polymorphism yields different results; for small N_S, $\hat{F}_{IS}$ is 0.061 ± 0.083 and for large N_S, $\hat{F}_{IS}$ is 0.107 ± 0.032. Unlike the *Es-3* polymorphism the *Hbb* polymorphism is very difficult to score accurately because the differences between the Hbb^dHbb^d and Hbb^dHbb^s phenotypes are very slight using the techniques applied by Selander *et al.* (1969). For this reason, the data on the *Es-3* locus are probably more reliable and the results obtained from the *Hbb* locus may be ignored as only a slight bias in scoring *Hbb* genotypes could explain the discrepancy.

The $\hat{F}_{IS}$ estimates obtained from the *Es-3* data are compatible with current ideas about house-mouse population structure (Anderson, 1970; Berry, 1970; Selander and Yang, 1970). If $\hat{F}_{IS(1)}$ is about -0.127 and $\hat{F}_{IS(\infty)}$ is about 0.052, then from Eq.(3.32B) and Eq. (3.34), N is about 6.5 and m is about 0.18. These estimates of N and m should be treated with scepticism as the theoretical model is only a simple approximation to a very complex real situation and the standard errors on the F_{IS} estimates are large.

3.12 DISCUSSION

The equations for F_{ST} (Eq.(3.20) and Eq. (3.29)) are different from the usual equations presented in the literature because the derivations do not presuppose that $F_{IS} = 0$ and the usual equations are applied to monoecious populations. The difference between monoecious and dioecious populations can be demonstrated by providing the appropriate equations for monoecious populations. The equivalent to Eq. (3.11) is

$$F_{IT}^{t} = F_{IT}^{t-1} + \frac{1}{2N_{t-1}}(1 - F_{IT}^{t-1})$$

Eq. (3.2) for F_{ST}^{t} will be the same, but if $F_{IT}^{o} = F_{ST}^{o} = 0$, then $F_{ST}^{(1)} = F_{IT}^{(2)} = \frac{1}{2N_1}$ and subsequently, $F_{IT}^{t} = F_{ST}^{t-1}$ so that the equivalent of Eq. (3.21) becomes

$$F_{ST}^{t} = F_{ST}^{t-1} + \frac{1}{2N_t}(1 - F_{ST}^{t-1}) \qquad \ldots\ldots(3.35)$$

The equation for $\hat{F}_{IS}$ (3.12) yields

$$\hat{F}_{IS}^{t} = \frac{-1}{2N_t - 1} \qquad \text{(cf. Crow and Kimura,1970)} \ \ldots\ldots(3.36)$$

In the case of a deme receiving immigrants, the equation for F_{IT}^{t} in monoecious populations is

$$F_{IT}^{t} = (1-m)^2\{F_{IT}^{t-1} + \frac{1}{2N_{t-1}}(1-F_{IT}^{t-1})\} \qquad \text{(Wright, 1969)}$$

and hence

$$F^{\infty}_{IT} = \frac{(1-m)^2}{2N-(2N-1)(1-m)^2}$$

The equation for F^{t}_{ST} will be the same as Eq.(3.29), but because of the alteration in F^{∞}_{IT}, the monoecious equivalent to Eq.(3.31) will be

$$F^{\infty}_{ST} = \frac{1}{2N-(2N-1)(1-m)^2} \qquad \ldots\ldots(3.37)$$

and from Eq.(3.12) the expression for $\hat{F}^{\infty}_{IS}$ is

$$\hat{F}^{\infty}_{IS} = -\frac{1}{2N-1} \qquad \ldots\ldots(3.38)$$

Unlike the dioecious situation, the value of $\hat{F}_{IS}$ for monoecious populations is independent of the immigration rate.

It is important to note that Eq.(3.35) and Eq.(3.37) can be obtained from first principles (cf. Eq.(3.18) and Eq.(3.27))with the assumption that $\hat{F}^{t*}_{IS_i} = 0$. Under these circumstances the monoecious equivalent of Eq.(3.29) is

$$F^{t}_{ST} = F^{t-1}_{ST}\left(1 - \frac{1}{2N_t}\right)(1-m)^2 + \frac{1}{2N_t} \qquad \text{(Wright,1969)}$$

There is an important difference between dioecious and monoecious populations in the behaviour of F_{IT}: in the first case, F_{IT} is zero for two generations after a small population is established from a large non-inbred population and in the second case F_{IT} is zero for only the first generation. The gap between F_{IT} and F_{ST} in the same generation produces the negative values of F_{IS}. However, in monoecious populations the gap disappears at the time when sampling occurs each generation because F_{IT} has advanced one generation and caught up with F_{ST} in magnitude. Therefore the equation for F^{t}_{ST} in monoecious populations need not include an F_{IT} term. In dioecious populations the equation for F^{t}_{ST} must be based on the assumption that F_{IS} is not zero and will include an F_{IT} term.

These comments imply that diffusion models (which always assume $F_{IS} = 0$) will be biased for dioecious populations.

The analysis of the sibmating case presented here differs from the analysis of Wright (1965, 1969) because Wright 'censuses' his population at a different stage of the life cycle. Wright's analysis is based on genotype frequencies after random union of gametes, i.e. in the zygotic genotypic array. This slight difference in timing leads to two changes in the results: (1) Wright finds that $F_{IT} = 0$ for only one generation and has only a one generation lag between F_{ST} and F_{IT}. This occurs because in the zygotic array F_{IT} has advanced a generation whilst F_{ST} remains at the value in the parental population. Hence in Wright's analysis, F_{IT} is one step ahead of its value in the analysis presented here. (2) When calculating the values of F_{IS_i} for the mating types Wright gives the value of zero to m_4 instead of -1. This is because he is calculating the F_{IS_i} in the progeny of each mating type, and not in the mating type itself. Consequently the estimate of $\hat{F}_{IS}$ that Wright obtains applies to the zygotic genotypic array (which cannot be censused in real populations) and not the adult genotypic array, which can be censused. The F-statistics used in this paper refer to the adult genotype distribution, before immigration (if it occurs) takes place. Hence if it is possible to census at this time, the results presented here will be directly applicable.

CHAPTER 4

ESTIMATION OF THE FIXATION INDEX IN A SUBPOPULATION

4.1 INTRODUCTION

Chapter 3 dealt with the problems of defining F_{IS} in the most useful way and the application of the mathematical theory of F_{IS} in a dioecious species to the interpretation of data from natural populations. However, several problems relating to F_{IS} remain to be solved. Firstly, in Chapter 3 the mathematical theory of F_{IS} was done indirectly by first finding equations for F_{ST} and F_{IT} and later substituting into Equation (3.12). Can a recurrence relation for F_{IS} be derived independently of F_{ST} and F_{IT}? Secondly, only brief attention was given to the change in F_{IS} that occurs when a sample is drawn from a population. What is the change in F_{IS} when a sample is drawn from a population? Thirdly, given a sample from a population, can the value of F_{IS} in the population be estimated? Fourthly, all the mathematical theory of F_{IS} so far has dealt with population with non-overlapping generations. Can similar results be obtained for F_{IS} in a species with overlapping generations?

In this chapter I will provide answers to the first three questions, but not the fourth. The analysis of inbreeding in populations with overlapping generations is very complicated (e.g. Felsenstein, 1971) and I have not made sufficient progress to be able to present any results here.

4.2 A RECURRENCE RELATION FOR F_{IS} IN AN ISOLATED DEME

Consider a dioecious species and the isolated deme model proposed in Chapter 3. Let the gene frequency in female gametes be $q_{f(t-1)}$ and in male gametes be $q_{m(t-1)}$, in generation (t-1). After random mating, the zygotic genotypic array is $q_{m(t-1)}\, q_{f(t-1)}AA$: $(q_{m(t-1)}(1-q_{f(t-1)}) + q_{f(t-1)}(1-q_{m(t-1)}))A\alpha$: $(1-q_{m(t-1)})(1-q_{f(t-1)})\alpha\alpha$. The mean gene frequency in generation t will be

$$\bar{q}_t = \tfrac{1}{2}(q_{m(t-1)} + q_{f(t-1)}) \qquad \ldots\ldots(4.1)$$

Let $$q_{m(t-1)} = \bar{q}_t - \delta q_{(t-1)} \qquad \ldots\ldots(4.2a)$$

and $$q_{f(t-1)} = \bar{q}_t + \delta q_{(t-1)} \qquad \ldots\ldots(4.2b)$$

where $$\delta q_{(t-1)} = \tfrac{1}{2}(q_{f(t-1)} - q_{m(t-1)}) \qquad \ldots\ldots(4.3)$$

is half the difference in gene frequencies between the two sexes at mating. The fixation index in the zygotic genotypic array (i.e. at gametic union) is

$$F_{IS}^{(t-1)*} = (\delta q_{(t-1)})^2/\bar{q}_t(1-\bar{q}_t) \qquad \ldots\ldots(4.4)$$

as was shown by Robertson, (1965).

The adults of the t-th generation are randomly chosen from this zygotic genotypic array with fixation index $F_{IS}^{(t-1)*}$. When N_f females are randomly chosen from the zygotic array, their gene frequency will be

$$q_{f(t)} = \bar{q}_t + \delta q_{ft} \qquad \ldots\ldots(4.5)$$

where $$E(\delta q_{ft}) = 0$$

and, from (3.16b),

$$E(\delta q_{ft})^2 = \bar{q}_t(1-\bar{q}_t)\ (1+F_{IS}^{(t-1)*})/2N_f \qquad \ldots\ldots(4.6)$$

Similarly, when N_m males are randomly chosen,

$$q_{m(t)} = \bar{q}_t + \delta q_{mt} \qquad \ldots\ldots(4.7)$$

and $$E(\delta q_{mt})^2 = \bar{q}_t(1-\bar{q}_t)(1+F_{IS}^{(t-1)*}/2N_m \qquad \ldots\ldots(4.8)$$

From Equations (4.3), (4.5) and (4.7),

$$\delta q_t = \tfrac{1}{2}(\delta q_{ft} - \delta q_{mt}) \qquad \ldots\ldots(4.9)$$

and therefore

$$E(\delta q_t)^2 = \bar{q}_t(1-\bar{q}_t)(1+F_{IS}^{(t-1)*})(1/8N_m + 1/8N_f) \qquad \ldots\ldots(4.10)$$

The $(1/8N_m + 1/8N_f)$ term can be replaced by $1/2N_e$, where N_e is the variance effective number of males and females. After random mating between the N_m males and N_f females, the zygotic array for the next generation will be formed with fixation index (from Equation (4.4)),

$$F_{IS}^{(t)*} = -(\delta q_t)^2 / \bar{q}_{(t+1)}(1-\bar{q}_{(t+1)}) \qquad \ldots\ldots(4.11)$$

However, this is the value of F_{IS}^{*} in one of the many possible replicate subpopulations that could have undergone the same sampling process in generation t. To obtain the expected value of $F_{IS}^{(t)*}$ over the ensemble of replicate subpopulations, Equation (3.11) (c.f. Cockerham, 1973) must be used:

$$E(F_{IS}^{(t)*}) = E(\bar{q}_{t+1}(1-\bar{q}_{t+1})F_{IS}^{(t)*})/E(\bar{q}_{t+1}(1-\bar{q}_{t+1})) \qquad \ldots\ldots(4.12)$$

$$= -E(\delta q_t)^2/E(\bar{q}_{t+1}(1-\bar{q}_{t+1})) \qquad \ldots\ldots(4.13)$$

since

$$E(\bar{q}_{t+1}(1-\bar{q}_{t+1})) = \bar{q}_t(1-\bar{q}_t)(1 - (1+F_{IS}^{(t-1)*})/2N_e) \qquad \ldots\ldots(4.14)$$

we find that, from Equations (4.10) and (4.14),

$$E(F_{IS}^{(t)*}) = -(1+F_{IS}^{(t-1)*})/(2N_e - (1+F_{IS}^{(t-1)*}))$$

If we now assume that $F_{IS}^{(t)*}$ is the expected value of F_{IS}^{*} in the ensemble of replicate demes in generation t, we have the required equation for $F_{IS}^{(t)*}$:

$$F_{IS}^{(t)*} = -(1+F_{IS}^{(t-1)*})/(2N_e - (1+F_{IS}^{(t-1)*})) \qquad \ldots\ldots(4.15)$$

F_{IS}^{*} soon reaches an equilibrium value if $2N_e$ remains constant. By substituting $F_{IS}^{(\infty)*}$ on both sides in Equation (4.15) at equilibrium we find

$$F_{IS}^{(\infty)*} = N_e - \sqrt{N_e^2 + 1} \qquad \ldots\ldots(4.16a)$$

$$\doteqdot -1/2N_e \qquad \ldots\ldots(4.16b)$$

Numerical checks indicate that Equation (4.16b) is a very good approximation to Equation (4.16a). This section has dealt with F_{IS} at the same stage of the life-cycle (gametic union) as Wright (1969,p.324). Substitution of $N_e = 2$ in Equation (4.15) and starting with $F_{IS}^{(0)*} = 0$, yields $F_{IS}^{(1)*} = -.333$; $F_{IS}^{(2)*} = -.200$; $F_{IS}^{(3)*} = -.250$; and $F_{IS}^{(\infty)*} = -.236$; as in Wright (1969), thus providing a check on this formula.

4.3 F_{IS} IN A SAMPLE.

Consider a population with fixation index F (for generality, the "IS" subscript will be dropped) and gene frequency $\bar{q}$. A sample of N_S diploids is drawn at random from the population, and the sample gene frequency and fixation index are q and F_S, respectively.

$$F_S = (2q(1-q)-H_o)/(2q(1-q)) \qquad \ldots\ldots(4.17)$$

where H_o is the frequency of heterozygotes in the sample.
From Equation (4.12),

$$E(F_S) = E(q(1-q)F_S)/E(q(1-q))$$

$$= (E(q(1-q)) - \tfrac{1}{2}E(H_o))/E(q(1-q)) \qquad \ldots\ldots(4.18)$$

As in Equation (4.14),

$$E(q(1-q)) = \bar{q}(1-\bar{q})(1 - (1+F)/2N_S) \qquad \ldots\ldots(4.19)$$

The frequency of heterozygotes in the population $(2\bar{q}(1-\bar{q})(1-F))$ is the value of $E(H_o)$ because, on average, there should be the same frequency of heterozygotes in a sample as in the population. Therefore, from Equations (4.18) and (4.19),

$$E(F_S) = ((2N_S-1)F - 1)/(2N_S - 1 - F) \qquad \ldots\ldots(4.20)$$

is the equation for the expected fixation index in a sample of N_S from a population with fixation index F.

In Chapter 3, the expected F_{IS} in adults was obtained whilst in section 4.2, the expected value at gametic union was derived. Since the adults are a sample from the zygotic array at gametic union, Equation (4.20) can be used to find the fixation index in N_S adults in an isolated deme:

$$F_{IS} = ((2N_S-1)F^*_{IS} - 1)/(2N_S-1-F^*_{IS}) \qquad \ldots\ldots(4.21)$$

This equation is quite general; it will apply whether or not N_S equals N_e and whether or not F^*_{IS} has reached equilibrium. As a check, when $F^*_{IS} = 0$ (as it does in monoecious populations), $F_{IS} = -1/(2N_S-1)$ (c.f. Equations (3.36) and (3.38)) in a sample of adults. If $N_S = N_e$ and N_e has remained constant for sufficiently long, then substituting Equation (4.16b) into Equation (4.21), yields

$$F^{\infty}_{IS} = -(4N_e-1)/(2N_e(2N_e-1)-1) \qquad \ldots\ldots(4.22a)$$

$$\doteqdot -1/N_e \qquad \ldots\ldots(4.22b)$$

Equation (4.22b) is in agreement with Equation (3.32b), and might have been deduced from it.

4.4 ESTIMATION OF F_{IS} IN A POPULATION

Two solutions have been proposed to the question: what equation is a suitable estimator of the fixation index in the population (assumed infinite) from which a sample with fixation index F_S was drawn? Neel and Ward (1972) using a formula due to Cannings and Edwards (1969), obtained the following estimator of a subpopulation F-statistic:

$$F_2 = ((2N_S-1)F_S + 1)/(2N_S+1-F_S) \qquad \ldots\ldots(4.23)$$

(I have rearranged their equation into the terms used in this chapter.) However, Cockerham (1973), suggested that

$$F_1 = ((2N_S-1)F_S+1)/(2N_S-1+F_S) \qquad \ldots\ldots(4.24)$$

TABLE 4.1 Computer simulation of sampling from nine genotypic arrays to test for bias in two estimators of the fixation index in a population.

	Gene frequency ($\bar{q}$)		
	.50	.45	.322
Population F	.0000	-.0101	-.0169
Sample F_S	-.0233 ± .0032	-.0342 ± .0032	-.0439 ± .0031
Estimator F_1	.0010 ± .0032	-.0099 ± .0032	-.0195 ± .0031
Estimator F_2	.0034 ± .0031	-.0069 ± .0031	-.0162 ± .0030
Population F	.4000	.3939	.2122
Sample F_S	.3755 ± .0030	.3709 ± .0030	.1779** ± .0033
Estimator F_1	.3962 ± .0029	.3915 ± .0030	.2013** ± .0032
Estimator F_2	.3861** ± .0029	.3816** ± .0029	.1957** ± .0031
Population F	-.4000	-.4141	-.2459
Sample F_S	-.4207 ± .0029	-.4322 ± .0028	-.2652 ± .0027
Estimator F_1	-.4004 ± .0029	-.4122 ± .0028	-.2421 ± .0027
Estimator F_2	-.3710** ± .0027	-.3818** ± .0026	-.2256** ± .0025

** $P < 1\%$ for Standard Normal Deviate test of appropriate null hypothesis: F_S tested against Equation (4.20); F_1 and F_2 tested against F.

Sample size = 20 diploids; 5,000 replicate samples from each genotypic array; results presented as mean ± SE.

was the appropriate estimator, and criticised the formula of Cannings and Edwards (1969) on which Equation (4.23) is based. F_1 and F_2 differ only in their denominators.

My approach to this problem is simple, but (it must be acknowledged) not very rigorous at present. By taking the observed F_S as $E(F_S)$ in Equation (4.20), that equation can be rearranged to find F in terms of N_S and F_S. It is found that the equation for F obtained in this manner is the same as Equation (4.24); that is, I agree with Cockerham (1973).

Computer simulation was used to determine if either of Equations (4.23) and (4.24) were biassed. A computer program was written to simulate the random sampling of N_S diploids from a genotypic array which represented a population with gene frequency $\bar{q}$ and fixation index, F. After each sample was drawn, the sample gene frequency (q), fixation index (F_S), F_1 and F_2 were calculated. When 5,000 independent samples had been drawn from a particular genotypic array, the mean and standard error of q, F_S, F_1 and F_2 were calculated. The results for samples of size $N_S = 20$, from nine genotypic arrays designed to cover a range of values of $\bar{q}$ and F are given in Table 4.1. Extreme values of F were chosen because a bias in F_1 or F_2 is more likely to show up under extreme conditions, even though such values are unlikely to occur in real populations. Each mean F_S was tested against the value predicted by Equation (4.20) and only one mean was significantly different from its expectation. The standard errors of F_S are in accord with the sampling variance formula of Rasmussen (1964):

$$\sigma_{F_S}^2 = (1-F_S)(2\bar{q}(1-\bar{q})(1+F_S) + F(2-F)(1-2\bar{q})^2)/2N_S\bar{q}(1-\bar{q}) \quad \ldots\ldots (4.25)$$

The only F_1 mean that was significantly different from F was in the same sampling experiment in which F_S was also significantly deviant and hence this single significant deviation of F_1 from F is probably due to

sampling error. F_2, on the other hand, is biassed for large positive or negative values of F because it is always highly significantly different from F in these cases. When F is near zero, there does not appear to be any difference between F_1 and F_2, but perhaps if the sampling experiments were repeated with more replicates, a difference might appear. Additional computer simulations with different N_S values and genotypic arrays confirmed that F_1 is apparently an unbiassed estimator of F (but note that computer simulation does not provide a complete proof of unbiassedness).

Another useful observation can be made from the computer simulation results. In all cases, the standard error of F_1 is the same or slightly smaller than the standard error of F_S. Hence Equation (4.25) also provides a good approximation to the sampling variance of F_1. This conclusion can be confirmed by noting that when $F<<(2N_S-1)$, then from Equation (4.24), $F_1 \doteqdot F_S + 1/(2N_S-1)$ and hence the variance of F_1 will be very similar to that of F_S.

So far, an effectively infinite population has been assumed. But many real populations are finite and a sample may include a sizeable fraction of the population. Under these circumstances, sampling without replacement implies that the hypergeometric distribution must be used instead of the binomial distribution when evaluating $E(\delta q)^2$. If the population contains T diploids, then

$$E(\delta q)^2 = \bar{q}(1-\bar{q})(1+F)A/2N_S \qquad \ldots\ldots(4.26)$$

where $A = (T-N_S)/(T-1)$ is the approximate correction for sampling without replacement when $T>>1$. From the methods used previously,

$$F_S = ((2N_S-A)F - A)/(2N_S-A(1+F)) \qquad \ldots\ldots(4.27)$$

and $$F_1 = (2N_SF_S + A(1-F_S))/(2N_S-A(1-F_S))$$

are appropriate when samples are drawn from populations that are not enormous.

4.5 DISCUSSION

This chapter provides the basis for a comprehensive theory of F_{IS} in subdivided populations when there are non-overlapping generations. The expected F^{*}_{IS} in a subpopulation receiving immigrants at a steady rate has not been dealt with, but this problem should be quite straightforward. The formulae given here would be of great value in analysing experiments such as that of Sing *et al.*, (1973) where many lines of *Drosophila* were set up with a known number of parents each generation. Each line is scored for several loci with codominant alleles so that it might be possible to test for heterogeneity of F_{IS} estimates between lines and between loci. If N_e is known, $F^{(t)}_{IS}$ can be calculated from Equation (4.21) and compared with the observed value. Alternatively, observed values of F_{IS} can be used to estimate N_e, although if there is heterotic selection, F_{IS} will be more negative and N_e will be underestimated. After several generations, ($t \geq 5$) if N_e can be kept fairly constant, then Equation (4.24) can be used to estimate $F^{(t)*}_{IS}$, and substitution of $F^{(t)*}_{IS}$ into Equation (4.16a) or (4.16b) will yield an estimate of N_e.

F_{IS} is a simpler F-statistic to use when estimating N_e in inbreeding lines because it stabilises in only 4 or 5 generations whereas F_{ST} and F_{IT} continue to change and a more complicated analysis is required. F_{ST} and F_{IT} can also be more difficult to calculate because it may be awkward to define the "total" population.

Cockerham (1973) has dealt with the problems of testing hypotheses about genotype frequencies in population samples. He recommends variance ratio and Chi-square tests, but Standard Normal Deviate tests based on Equation (4.25) could also be used to test any appropriate null hypothesis about F_{IS}.

Populations with overlapping generations create a whole new set of problems. The presence of several age-classes not only intermingles individuals with a greater variety of genetical relationships (e.g. parent-offspring) but also effectively further subdivides each subpopulation because each age-class can have different mortality and reproduction characteristics. My view, at present, is that none of the formulae given in Chapters 3 or 4 are entirely satisfactory for populations with overlapping generations. These equations suggest what might happen in subdivided populations with overlapping generations but do not provide valid estimates of F_{IS} in such populations. The little work that I have done (mainly computer simulation) indicates that F_{IS} may be even more negative when generations overlap, but because this work has not extended sufficiently far for me to have confidence in it, I will not present any of it.

CHAPTER 5

A RELATIONSHIP BETWEEN TWO METHODS FOR STUDYING INBREEDING IN SUBDIVIDED POPULATIONS

5.1 INTRODUCTION

The mathematical techniques for studying subdivided populations known as the F-statistics and as the probability of identity methods, were introduced in Chapter 2, where it was noted that F_{IT} can be equated to the probability of identity by descent of uniting gametes. "Probability of identity" theory is no longer restricted to probability of identity by descent but also includes probability of identity in state (e.g. Maruyama, 1972). Cockerham (1973) and Nei (1973) have recently expressed a relationship between two measures of the probability of identity of two randomly chosen alleles and F_{ST}. Because this relationship bridges the gap between the F-statistics mathematical theory and probabilities of identity mathematical theory, I will present my approach to this topic in this chapter (this was evolved independently of Nei and Cockerham).

5.2 F-STATISTICS FOR MULTIPLE ALLELES

Nei (1965) attempted to generalise the F-statistics to a multi-allelic locus, but because his definition of F_{IS} was poorly chosen (Chapter 3), the remainder of his results are suspect. Li (1969) proposed several indices to measure the association between alleles within subdivisions for a multi-allelic locus, but was unable to decide which index was the best generalisation of F_{ST}. In this section, I will derive an equation for F_{ST} at a multi-allelic locus in a subdivided population that satisfies the relationship between F-statistics,

$$(1-F_{IT}) = (1-F_{IS})(1-F_{ST}) \qquad \ldots\ldots(5.1)$$

and the definition of F_{IS} (from Equation (3.11))

$$F_{IS} = (\Sigma w_i\, q_i(1-q_i)F_{IS_i})/(\Sigma w_i\, q_i(1-q_i)) \qquad \ldots\ldots(5.2)$$

For a locus with ℓ alleles, the frequencies can be denoted $(q_{1i}, q_{2i}, q_{3i} \ldots q_{\ell i})$ in the ith subpopulation and $(q_1, q_2, q_3 \ldots q_\ell)$ in the total population.

Hence $$q_l = \sum_i w_i q_{li} \qquad \ldots\ldots(5.3)$$

where $\sum_i$ indicates summation over all subpopulations, and w_i is the proportion of the total in the ith subpopulation.

I will begin by defining a number of relationships that are needed later. The covariance $(\sigma_{q_j q_k})$ between the jth and kth alleles over all subpopulations is

$$\sigma_{q_j q_k} = \sum_i w_i q_{ji} q_{ki} - q_j q_k \qquad \ldots\ldots(5.4)$$

The frequency of heterozygotes in the total population (H_T) will be $(1-F_{IT})$ times the frequency expected by the Hardy-Weinberg law:

$$H_T = \sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k \neq j}}^{\ell} q_j q_k \,(1-F_{IT}) \qquad \ldots\ldots(5.5)$$

Similarly, the frequency of heterozygotes in the ith subpopulation (H_i) is

$$H_i = \sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k \neq j}}^{\ell} q_j q_k \,(1-F_{IS_i}) \qquad \ldots\ldots(5.6)$$

The appropriate generalisation of the definition for F_{IS} (Equation (5.2)) is

$$F_{IS} = \left(\sum_i w_i \sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k \neq j}}^{\ell} q_{ji}\, q_{ki}\, F_{IS_i}\right) \Big/ \left(\sum_i w_i \sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k \neq j}}^{\ell} q_{ji}\, q_{ki}\right) \qquad \ldots\ldots(5.7)$$

The total frequency of heterozygotes is also

$$H_T = \sum_i w_i H_i \qquad \ldots\ldots(5.8a)$$

Therefore, from equations (5.4), (5.6) and (5.7),

$$H_T = (1-F_{IS}) \sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k\neq j}}^{\ell} (q_j q_k + \sigma_{q_j q_k}) \qquad \ldots\ldots(5.8b)$$

Comparison of Equations (5.1), (5.5) and (5.8b) suggests that the appropriate definition of F_{ST} for a locus with multiple alleles is

$$F_{ST} = \left(\sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k\neq j}}^{\ell} \sigma_{q_j q_k} \right) \Big/ \left(\sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k\neq j}}^{\ell} q_j q_k \right) \qquad \ldots\ldots(5.9)$$

Equation (5.9) may be simplified by noting that

$$\sigma^2_{q_j} = - \sum_{\substack{k=1 \\ k\neq j}}^{\ell} \sigma_{q_j q_k} \qquad \text{(Li,1969)} \qquad \ldots\ldots(5.10)$$

$$\text{and } q_j(1-q_j) = q_j \sum_{\substack{k=1 \\ k\neq j}}^{\ell} q_k \qquad \ldots\ldots(5.11)$$

and hence,

$$F_{ST} = \left(\sum_{j=1}^{\ell} \sigma^2_{q_j} \right) \Big/ \left(\sum_{j=1}^{\ell} q_j(1-q_j) \right) \qquad \ldots\ldots(5.12)$$

This definition was briefly mentioned in Li (1969) and has also been suggested by Wright (quoted in Nei (1973)). Equation (5.12) shows that for the multi-allelic case, F_{ST} is still the ratio of the actual variance in gene frequencies over subdivisions to the limiting variance when all subpopulations are fixed for one allele or another. Another way of expressing Equation (5.12) is

$$F_{ST} = E(\sigma^2_q) / E(q(1-q))$$

where the expectation, $E(\cdot)$, is taken over all alleles at a locus.

Equations (5.9) and (5.11) show that there is only one value of F_{ST} per locus, regardless of the number of alleles (unless one chooses to define separate F-statistics for every pair of alleles, q_i and q_j, but this is very tedious and would be useful only in very special cases). This conclusion differs from that of Nei (1965), who could not decide

whether F_{ST} was $\sigma^2_{q_j}/q_j(1-q_j)$ or $-\sigma_{q_j q_k}/q_j q_k$. Because of this, F_{ST} could be estimated in $\ell(\ell+1)/2$ different ways for a locus with ℓ alleles, but because these various estimates of F_{ST} are not independent this can cause some difficulties (e.g. Lewontin and Krakauer, 1973). Another possible error is that the average of all these estimates of F_{ST} will not be the same as F_{ST} calculated from Equation (5.12).

5.3 THE PROBABILITY OF IDENTITY OF ALLELES

The probability of identity of two randomly chosen alleles is obviously $\sum_{j=1}^{\ell} q_j^2$. If each allele has arisen only once by mutation, or if each allele was present only once when the population was initiated, then $(\sum_{j=1}^{\ell} q_j^2)$ can be interpreted as the probability of identity by descent. But this term is the probability of identity in state (or function) if we know only that the alleles differ in state (or function) and similarity of state does not necessarily imply identity by descent. In a subdivided population, the probability of identity of two alleles, each randomly chosen from a different subpopulation, (say, the hth and the ith) is

$$f_{hi} = \sum_{j=1}^{\ell} q_{jh}\, q_{ji} \qquad \ldots\ldots(5.13)$$

Similarly,

$$f_{ii} = \sum_{j=1}^{\ell} q_{ji}^2 \qquad \ldots\ldots(5.14a)$$

$$= 1 - \sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k\neq j}}^{\ell} q_{ji}\, q_{ki} \qquad \ldots\ldots(5.14b)$$

is the probability of identity of two alleles randomly chosen from the ith subpopulation.

There are two ways of randomly drawing two alleles from the total population that are useful to consider: (a) a subpopulation may be chosen randomly, and the two alleles chosen randomly from within it (denoted f_o); or (b) the two alleles may be chosen completely at random from the total

population (denoted $\bar{f}$).

By definition, $f_o = \sum_i w_i f_{ii}$ .....(5.15a)

$= \sum_i w_i \sum_{j=1}^{\ell} q_{ji}^2$ .....(5.15b)

$= 1 - \sum_i w_i \sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k \neq j}}^{\ell} q_{ji} q_{ki}$ .....(5.15c)

and similarly, $\bar{f} = \sum_h w_h \sum_i w_i f_{hi}$ .....(5.16a)

$= \sum_{j=1}^{\ell} q_j^2$ .....(5.16b)

$= 1 - \sum_{j=1}^{\ell} \sum_{\substack{k=1 \\ k \neq j}}^{\ell} q_j q_k$ .....(5.16c)

The magnitude of f_o will be less than that of $\bar{f}$ if the population is subdivided and gene frequencies differ between subdivisions.

5.4 THE RELATIONSHIP BETWEEN F_{ST}, f_o AND $\bar{f}$

The ratio, $(1-f_o)/(1-\bar{f})$, has been proposed as a measure of the differentiation of gene frequencies between subpopulations (Kimura and Maruyama, 1971; Maruyama, 1972). Because F_{ST} is also proportional to the amount of variation between subpopulations, a reasonable assumption is that the above ratio and F_{ST} ought to be related.

From Equations (5.6) and (5.14b),

$$H_i = (1-f_{ii})(1-F_{IS_i}) \qquad \text{.....(5.17)}$$

then from Equations (5.8a), (5.7) and (5.15c),

$$H_T = (1-f_o)(1-F_{IS}) \qquad \text{.....(5.18)}$$

and from Equations (5.5) and (5.16c),

$$H_T = (1-\bar{f})(1-F_{IT}) \qquad \ldots\ldots(5.19)$$

Inspection of Equations (5.18) and (5.19) shows that

$$(1-f_o)/(1-\bar{f}) = (1-F_{IT})/(1-F_{IS})$$

$$= 1 - F_{ST} \qquad \ldots\ldots(5.20a)$$

$$\text{or,} \qquad F_{ST} = (f_o-\bar{f})/(1-\bar{f}) \qquad \ldots\ldots(5.20b)$$

Equation (5.20b) has been obtained by Cockerham (1973) and Nei (1973), although Nei denoted F_{ST} by G_{ST} because of his different definition for F_{ST}.

Equation (5.20a) can be obtained very simply for the case of a di-allelic locus by calculating the variance effective number of a subdivided population with n subdivisions, each of variance effective size N. If each subdivision has Hardy-Weinberg frequencies at gametic union, then the variance effective number of the total population is from Wright (1943a),

$$N_e = nN/(1-F_{ST}) \qquad \ldots\ldots(5.21a)$$

or from Maruyama (1972),

$$N_e = nN(1-\bar{f})/(1-f_o) \qquad \ldots\ldots(5.21b)$$

Comparison of Equations (5.21a) and (5.21b) yields (5.20a).

Perhaps it ought to be mentioned that $f_o \neq F_{IT}$ and $\bar{f} \neq F_{IS}$, although these equations might appear obvious from Equation (5.20a). F_{IT} is the probability of identity of uniting gametes in the whole population while f_o is the average probability of identity of two random alleles within subpopulations. F_{IS} is often negative within subdivisions so it cannot be the same as $\bar{f}$, which cannot be negative.

From Equation (5.20b) it can be confirmed that F_{ST} is a fixation index.

$$\text{Let} \qquad \bar{h} = 1 - \bar{f} \qquad \ldots\ldots(5.22)$$

be the frequency of heterozygotes that are expected in the population in the abscence of population subdivision.

Also, $h_o = 1 - f_o$ .....(5.23)

is the frequency of heterozygotes that should be observed taking account of population subdivision but neglecting the effects of deviations from Hardy-Weinberg frequencies within subpopulations. The F-statistics can now be arranged:

From Equation (5.18), $F_{IS} = (h_o - H_T)/h_o$ .....(5.24a)

from Equation (5.20b), $F_{ST} = (\bar{h} - h_o)/\bar{h}$ .....(5.24b)

and from Equation (5.19) $F_{IT} = (\bar{h} - H_T)/\bar{h}$ .....(5.24c)

Whereas F_{IS} and F_{IT} are deviations between expected and observed frequencies of heterozygotes, F_{ST} is the deviation between the expected frequency of heterozygotes with population subdivision and the expected frequency without subdivision. F_{ST} measures the amount of fixation due to population subdivision.

5.5 DISCUSSION

This chapter has shown that at least two of the results, obtained by F-statistics methods and by probability of identity methods, are the same. However, this is only the beginning of an attempt to bring together these two methods of dealing with subdivided populations. For example, the relationship between F_{IS}, F_{ST} F_{IT}, f_o, $\bar{f}$, C(r) and f(r) (see Chapter 2) in isolation by distance and stepping-stone models of population structure are not fully understood and this could be a profitable area of research.

Nei (1973) derived a new measure of "gene diversity" (denoted D_{ST}) between subpopulations. Comparison of his results with mine (especially his Equation (9) with Equation (5.12)) suggests that

$$D_{ST} = \sum_{j=1}^{\ell} \sigma^2_{q_j}$$

His method of partitioning the gene diversity within and between sub-populations appears to simply be a partitioning of variances. Cockerham (1969,1973) has already dealt with the analysis of variance approach to subdivided populations.

CHAPTER 6

HOUSE MOUSE ECOLOGICAL GENETICS

6.1 INTRODUCTION

The house mouse (*Mus musculus*) is a popular animal for ecological and genetical research. This species is widespread around the world (Berry, 1970; Schwarz and Schwarz, 1943); occupies a wide range of habitats (e.g. houses, barns, cold stores, farmlands, haystacks; Southern, 1954); is often the most common mammal around human habitation; is an economically important pest (Southern, 1954); and has many tame laboratory strains (usually highly inbred) that facilitate experimental work. Mice are small so that large numbers are easily maintained, easy to keep and breed readily with large litters at monthly intervals.

Most of the studies on mouse populations have been primarily ecological or genetical. I will review the ecological and genetical studies separately and attempt a synthesis at the end.

Wild mice are small (usually less than 20 gm, Berry, 1970), an advantage when seeking cover from predators or inclement weather because a relatively small hole can shelter a mouse, but a disadvantage in extreme temperatures because of the high surface area to volume ratio (Southern, 1954). This disadvantage, however, has not prevented mice from living in cold stores at -10^{o}C to -20^{o}C when they can obtain insulated bedding (feathers, fur, hessian bagging) and eat highly nutritious foods (Laurie, 1946), or from living in semi-arid areas of South Australia where daytime temperatures exceed 40^{o}C in the shade, provided they can burrow underground to keep cool and conserve moisture (Newsome, 1969a,b; P. Aitken, personal communication). Thus, by behaviour suited to the environment, mice can live in apparently adverse climates.

Mice are omnivores. Southern (1954) observed that cereals of all kinds are the staple diet of wild mice and that meat is utilised in cold

stores whilst in dry environments succulent foods may be nibbled for the water they contain. Whitaker (1966) in a study of the stomach contents of 478 house mice in Indiana confirmed that seeds and cereals were the most important dietary component, and found that animals (especially insect larvae and pupae) provided a major source of food because about 70% of stomachs contained some animal remains. Berry (1968) reported that mice on Skokholm Island were also omniverous: nearly all stomachs contained plant material and over 50% contained arthropod remains.

Because mice can tolerate such wide variations in climate and food supply, they can live in a wide range of habitats. The many studies of house mouse ecology cover this range of habitats, so to simplify this review, I will broadly group the studies according to the habitat in which the mice were living. This may appear arbitrary, but this way of classifying different studies is useful because mouse populations in similar habitats in different parts of the world are more alike ecologically than mouse populations living in different habitats in the same locality.

6.2 ECOLOGY OF THE HOUSE MOUSE

6.2.1 Commensal Mice in Buildings

Laurie (1946) reported on studies of house mice in England carried out with the intention of developing better methods of reducing the damage done to foodstuffs by commensal mice. Three habitats in buildings were considered in particular: cold stores, buffer depots and urban environments (ordinary warehouses, shops, restaurants and houses). In cold stores and buffer depots (where sacks of sugar or white flour were stored in large stacks), mice thrive because they live in the midst of food and are sheltered from predation or disturbances. But in urban environments, shelter is often separate from food, so that mice are obliged to run out into the open to obtain food and expose themselves to predators or traps. Food in urban sites is varied and sometimes

limited in quantity. Reproduction, as measured by the estimated number of female embryos produced per adult female per year, was least in urban environments, being only 2/3rds the value observed in cold stores and buffer depots.

Southern (1954) in a review of the house mouse research carried out in England during the Second World War, suggested that the major factors affecting mouse population dynamics are as follows.

(a) Proximity of cover and food supplies. Mice are most vulnerable to predators, traps or poisons when moving between home and food sources and consequently thrive best when food and shelter are combined. The "home range" of a mouse will depend upon the proximity of food sources. Experiments in a large room indicated that when there were many sources of food, the area covered by a mouse in its daily feeding and exploratory activities was only about 5 square metres. But when there was only a single food source, home ranges rose to about 20 square metres. Observations on a stack of grain sacks suggested that mice living up to six feet inside the stack often came to the periphery but no precise estimates of home ranges can be made under these circumstances because the mice are mostly inaccessible.

(b) Climate and food requirements. As previously mentioned, mice have a small bulk which makes them vulnerable to temperature extremes, although they often avoid the direct effects of extreme temperature by building insulated nests or burrows. Food consumption drops in hot weather although thirst increases, naturally. Southern and co-workers were able to keep wild mice on grain diets without water, but the mice did better when given water. Experiments with different diets suggest that grain alone is not optimal and leads to frequent eating of litters.

(c) Disturbance and nocturnal habits. Mice which are undisturbed have unlimited time for foraging and feeding, but when there is continual change in the physical environment (for example, with the rotation of stores in

large grocer's shops) mice will not settle down and breed. The basic feeding rhythm (with most feeding occurring at night) can be modified to suit the circumstances and daytime feeding can occur if conditions permit.

Southern (1954) also dealt with the problem of immigration of mice into areas that had been cleared of mice by trapping and poisoning campaigns. In shops, the effects of mouse destruction were sometimes quickly nullified by a high rate of immigration from neighboring buildings. After a warehouse had been cleared of mice, their spread was monitored after they re-entered. The mice moved 20 to 24 metres in 16 days, an average of at least 1.25 metres per day. The speed of advance was probably limited by the slowness with which mice explore unfamiliar territory. These observations suggest there is some form of internal pressure causing mouse populations to expand and occupy all the available territory that will support mice.

Brown (1953) trapped for two years in a barn on a horse breeding farm in Maryland. He deduced fluctuations in population density from changes in the trap rate (mice caught/100 hours traps open). The maximum trap rate occurred in late Autumn (November-December) and the minimum in Summer (July-August). Daily movements of mice and home ranges were estimated by measuring the distance between successive captures of each mouse. The average distance between recaptures was only 5.25 metres and home ranges were estimated at 6 metres across for males and only 3.8 metres across for females. Brown noted that male mice appeared to shift the centres of their home ranges with time but females remained at the same location. Male mice disappeared faster (died or emigrated) than females: estimated monthly mortalities (including emigration) were 0.43 for males and 0.24 for females. The higher mortality rate for males makes it seem surprising that there was an excess of males (63% males) among the mature mice that Brown caught. Since the sex ratio in immature mice was 1:1, these observations suggest that there was a higher immigration

rate for males than females, to more than compensate for the higher rate of loss of males. Laurie (1946), in contrast, found 50.07% males in her commensal populations.

Evans (1949) made a mark-recapture study of mice in a seed barn and neighboring outdoor areas, at Davis, California. He found that the indoor population was dense in winter but became extinct in summer until recolonisation occurred in autumn (November-December). He observed three mice to move from indoors to outdoors in spring and early summer and one outdoor to indoor movement in autumn.

The studies of Brown (1953) and Evans (1949) suggest that a barn in a farming area is subject to a considerable ebb and flow of mice. It is a refuge in winter after deterioration of the outdoor environment drives mice indoors in autumn. In spring, mice leave the barn to colonise the outdoor habitats again.

Smith (1954) trapped in a variety of indoor locations in Mississippi for a year. He obtained the highest trap rates in food handling premises and the lowest trap rates in farm buildings (where he notes there is an abundance of food, many predators and unspecified "other factors" that might reduce trap rates). He observed a 1:1 sex ratio overall, but there was a preponderance of females in late summer and a deficiency in early winter. Breeding occurred all year (Laurie (1946) also found this) with peaks in spring and autumn suggesting that the hot weather in mid-summer and the cold weather of winter inhibited reproduction. He found fewer embryos per pregnant female (all averages 5 or less) than Laurie (1946) (all averages greater than 5).

6.2.2 Wild Mice in Grasslands and Farmlands

Baker (1946) reported a study of rodents on partly cleared land on Guam Island in the Pacific Ocean. The estimated population density of house mice varied between 8/hectare and 25/hectare. 80% of recaptured

mice moved less than 23 metres between captures. There was a surplus of males.

Southern and Laurie (1946) trapped for small mammals in farmlands in the vicinity of corn-ricks (haystacks) in England. The house mouse was the third most common mammal, totalling about 20% of the catch. In summer, mice were caught in standing corn (wheat and oats), under stooks (sheaves piled together to dry), around ricks and in hedgerows. In winter they were restricted to ricks (discussed in 6.2.3) and hedgerows. They were never found in open fields until the crops had grown up sufficiently to provide cover. There was little or no breeding in winter outside ricks.

Evans (1949), also found a small outdoor population maintained all year with movement between the outdoor and a nearby indoor population.

Mice were caught at five grassland locations near San Francisco by Breakey (1963). He found a distinct breeding season (mid-April to November) with no mice born in winter and early spring. Few mice lived past eighteen months. Most mice that were mature in spring did not survive until the end of summer; consequently in autumn the average age of trappable mice was about three months. He observed that the age of sexual maturity depended upon season, because early in the breeding season two month old mice were breeding but in autumn the onset of fertility of young mice was delayed, often until the end of winter. Some males were fertile all year but females were unable to become pregnant for about four months of the year.

Pearson (1963) trapped house mice at two grassy sites near Berkeley, California, where they are not normally common. In both places, population densities reached 500-750 per hectare and remained high for slightly more than a year. There was considerable variation in population densities at different localities at the same time.

Lidicker (1966) observed the decline to extinction of a mouse population on a 22.3 hectare island in San Francisco Bay. Initially, in the autumn, the population density was about 750/hectare, but steadily declined for the next year until no mice were caught. This decline in the house mouse population coincided with the colonisation of the island by *Microtus californicus* , which appeared to supplant the mice. Home ranges of house mice averaged 145 square metres (the average maximum distance across each home range was 17.7 metres) with considerable overlap in home ranges of some mice. Females in summer had home ranges about half the average size, perhaps because of nesting and care of young. No evidence of territoriality could be found.

DeLong (1967) studied six mouse populations in old-field grasslands near San Francisco Bay for two years. Population densities fluctuated between less than 50/hectare and nearly 750/hectare. He proposed three explanations for the four major declines in population density that were observed: (a) epidemic disease which reduced one population from 500/hectare to about 60/hectare in one month; (b) insufficient food in the autumn; (c) "intrinsic mechanisms" reducing the number of weaned young per lactating female and cessation of breeding by females when food shortages were overcome by supplementary feeding and population densities remained high. Increased emigration and mortality rates for juveniles and sub-adults were observed in the two populations with supplementary feeding. Some of the dispersing mice were caught 400 metres from the study areas. Besides this dispersal, which appeared to be related to high population densities, a marked increase in the frequency of long range dispersal movements was observed on most study areas in autumn (October-November).

The average distance between first capture as a juvenile and last capture as an adult was larger during the breeding season (about 24 metres) than during the non-breeding season. The average distance between

successive captures within a trapping period (daily movements) was slightly greater in juveniles or subadults than adults.

The proportion of lactating adult females declined to zero in winter except when supplementary food was provided. Body growth rates declined in winter, probably because of lower temperatures and food shortages.

In another study near San Francisco Bay, Quadagno (1968)measured home ranges of house mice at two sites. On the first site,a population of *Microtus californicus* increased fourfold while the mouse population dwindled from 76/hectare in February to apparent extinction in April. The other site did not have a *Microtus* population and the mice increased from 100/hectare in September to 153/hectare in October. On the first site the average distance between recaptures was 6 metres, and more than twice this on the second site. Quadagno suggested that the *Microtus* on the first site restricted the movement of mice, but since the two sites differed in population densities and population dynamics as well, this suggestion is speculative.

Berry (1968) reported on his studies of wild mice on the 100 hectare island of Skokholm, 3 km off the coast of Pembrokeshire. The mice persisted all year on the coastal cliffs and colonised the interior of the island only in spring and summer. Colonisation of the interior was carried out by younger (smaller) mice in the springtime. Home ranges of established mice often overlapped, but there was little evidence of joint territoriality by pairs of adult mice. Considerable dispersal movements were observed in spring when many mice moved more than 180 metres and some even travelled the length of the island (1.5 Km). Estimates of the population size range from a low of 100-300 to a high of 2,500-5,000 (i.e. average density, 25-50/hectare). There was a heavy winter mortality followed by a 10-fold increase in numbers each summer. The breeding season began about mid-April and ended by October.

Newsome (1969a,b) studied two sites at Turretfield Agricultural Research Centre about 48 Km north-east of Adelaide in South Australia. One site was in a wheatfield and the other was a reed-bed in nearby Salt Creek. Mice inhabited the wheatfield for only a part of each year because winter rains water-logged the soil, making it uninhabitable (Newsome, 1969a). In spring and summer, emigrants from more permanent populations explored the wheatfield. If the summer were dry, these emigrants could not dig burrows in the hard soil and so could not establish themselves in the wheatfield. But if summer rains softened the soil and opened up cracks, mice were able to burrow down the cracks and obtain shelter from heat and dryness. Once resident in the wheatfield, the mice bred until they ran out of food, were flooded by winter rains or disturbed by ploughing. In localities where food was plentiful, and the mice became established sufficiently early in the summer, population densities soared until local "mouse plagues" formed (Newsome and Crowcroft, 1971).

The reed-bed lay about 180 metres downhill of the wheatfield and contained mice all year (Newsome, 1969b). This population was a reservoir from which the marginal habitat in the wheatfields was colonised each spring and summer. The reed-bed population increased each summer, regulated by the availability of burrowing sites and food supplies. In times of shortages of resources (space or food), Newsome suggested that social factors, in particular, dominance hierarchies, helped determine which mice were able to survive and which were forced out, usually to die. Smaller adult mice were the first to stop breeding when the environment deteriorated in winter and the last to resume breeding when spring came. In spring, smaller adults tended to disappear from the reed-bed and reappear as colonists in the wheatfield. In autumn and winter another movement of mice occurred, this time the smaller adults remained in the wheatfield when the environment deteriorated while larger mice emigrated and often

appeared in the reed-bed. Thus there was a seasonal flow of mice in and out of the reed-bed. Reproduction ceased each winter because females were anoestrous for 2 to 5 months.

In the wheatfield, population densities reached 100/hectare in 1965 and probably approached 1500/hectare in the 1970 mouse plague (which occurred after mice overwintered in the wheatfield in 1969 and subsequent conditions were ideal for population growth (Newsome, personal communication)). In 1965, the reed-bed population reached 600/hectare (Newsome, 1969b); the 1970 density is unknown but was probably similar to that in the wheatfield.

Dispersal movements were observed between several sites where trapping was carried out. Journeys between the wheatfield and creek were recorded and four mice moved over 250 metres between recaptures. One pregnant female moved from the creek to the wheatfield (183 metres) one night and back the next night. Newsome also measured the maximum distance between captures of a mouse in each trapping period (the range length). The average range length varied between 10 and 20 metres, but the only consistent change in both sexes was a marked reduction in average range length when population densities were high in 1965.

6.2.3 Wild Mice in Haystacks (or corn-ricks)

Laurie (1946) studied corn-ricks (haystacks) of wheat or oats within 30 miles of Oxford. She found that mice, living on the grain in the ricks, had unlimited food and shelter combined. The climate was ameliorated because temperature variations were less within a rick than outside. Mice lived in all layers of the ricks, even burrowing into the earth and decaying matter underneath, although invasion of each rick was thought to begin at the top. Most ricks were only temporary environments because they were built in August-September and destroyed for threshing in late winter and spring. In comparison with mice living in buildings, rick mice were more fertile and had a higher reproductive rate.

Southern and Laurie (1946) extended these observations and compared the rick mice with those living in the surrounding farmlands. In the summer, when few ricks are left standing, most of the mice are in sparse colonies in and around the cereal crops. In autumn, the crops are harvested and the mice are either carried (in sheaves) to the ricks or immigrate of their own accord. In winter, only a few mice remain in hedgerows, but large populations are rapidly breeding up in the ricks. These rick populations are dispersed at threshing time, most are killed but a few survive to recolonise the fields next summer.

By observing the destruction of many ricks, Southern and Laurie (1946) found that population densities ranged from 1 to 12.63 active mice/cu. metre (or up to 15/cu.m. if babies were counted). The proportion of ricks with more than 50 active mice was more than 50% by February (winter) each year. Rick populations reached maximum numbers 8 to 11 months after establishment.

Southwick (1958) also studied English ricks. He found a great variation in population densities of ricks 8-10 months old (2-20 mice/cu.m., or up to 16.18 weaned mice/cu.m.). Wheat ricks contained more mice than oat ricks and ricks near farm buildings had more mice than ricks further out in the fields. There was a small decline in fertility in the more densely populated ricks: pregnancy rate declined from 50% to 43%; average litter size fell from 6.23 to 5.11; embryonic and foetal resorption increased from 14% to 27% and the incidence of male mice with wounds increased (0.8% to 21.1%) in high density ricks as compared with low density ricks. But the reproductive performance of mice in overcrowded ricks was still better than that of many other mouse populations.

Measurements of the movement of mice in and out of four ricks were made by Rowe *et al.* (1963). Their ricks were constructed in August and the greatest influx of immigrants occurred in October, November and December, although some immigration occurred at all times. Emigration from the ricks

began in December and increased steadily until May, when the ricks were threshed. There were more males than females in the migrants in both directions with a slight female surplus within ricks at threshing. Rowe *et al.*(1964) also examined the effects of crowding within ricks on reproduction. The proportion of mice capable of breeding and litter size were higher in the less crowded ricks. They concluded that the rick populations were only slightly regulated by lowered fecundity, reduced litter size, embryonic resorption and emigration.

Newsome (1971) dealt with the population dynamics of mice in two kinds of haystacks in South Australia. Initially, he studied the numbers of mice caught in large stacks, built at chaff-mills, as the stacks aged. By fitting curvilinear regressions to his data, he obtained equations for the numbers of mice at different times after the stacks were built in the top, middle and bottom thirds of stacks made from oats, oats and wheat, or wheat only. Unlike Southwick (1958), Newsome found that oat stacks contained more mice than wheat stacks, particularly in the top third. In all stacks the top third contained the most mice and the middle third the least mice, probably because compression of the stack prevented ingress to the middle regions. The bottom of a stack was colonised by mice burrowing in the softened damp soil underneath. In all stacks, the population density rose rapidly for the first six months before peaking at 6 to 8 months and declining almost to zero in three years. Food shortages, due to the difficulties that the mice faced when burrowing through the compacted inner regions of the stacks to find unexploited grain, caused the population declines. The maximum density of mice was only 2.1/cu.m., much less than in English ricks, but there were signs of social disruption due to overcrowding such as wounds, disease and curtailment of breeding activities. The average density of mice may seem small, but they were not randomly distributed, being concentrated in tunnels and, presumably, near food sources.

Newsome (1971) also built a special small stack which he censused part by part at three monthly intervals. The population size peaked at six months and was reduced to one mouse after 15 months. The maximum density was 12.1/cu.m., when there was evidence of neglect of young, cannibalism and fighting. However, this peak density was perhaps inflated by immigration from nearby fields that had been soaked by winter rains or ploughed. The oat stems in this stack were more loosely packed than in the large stacks, and this looseness probably permitted easier access to more of the grain in this stack than in the compacted stacks. The mouse population could grow more rapidly and support a greater total, but when food shortage started, it was due to an absolute lack of food rather than increasing difficulties associated with food gathering, and consequently, the population declined rapidly. Water shortages in the summer also affected the mice in this oat stack because they congregated in the moister regions.

6.2.4 Wild and Domestic Mice Confined in Pens

The behaviour of wild mice in their natural state is difficult to observe. Consequently, mouse behaviour and ecology have been studied on mice that are confined in pens. But we must be careful when extrapolating from studies on confined mice (which cannot emigrate) to mice living in the wild, where emigration from unfavourable localities is a normal and common event (see above review). Some behaviour patterns observed in confined mice may be quite abnormal, but we do not know for certain because comparable studies on unconfined mice have not been done. For the present, conclusions drawn from studies on penned mice should be regarded as provisional only.

When several mice are first introduced into a pen (which may be as small as three ordinary mouse cages joined together (e.g. DeFries and McClearn, 1970) or as large as a room (e.g. Crowcroft and Rowe, 1963)) there is a short period of exploratory behaviour in which each mouse becomes accustomed to the new surroundings. When two strange mice meet

for the first time, both will usually retreat (Crowcroft and Rowe, 1963; Reimer and Petras, 1967). Subsequently, fighting breaks out between males until a dominance hierarchy develops or all males but one are dead (Brown, 1953; Southwick, 1955a; Crowcroft,1955; Crowcroft and Rowe, 1963; Reimer and Petras, 1967; Oakeshott, 1974). Crowcroft and Rowe (1963) suggested that the area of a pen controlled by a dominant male depends upon the amount of cover present. If there is little or no cover, a single male can be dominant over all other males, even in a 6m. x 6m. room. But if the physical environment is sufficiently complex, with cover for mice to shelter in and obstacles to create natural boundaries within the pen, several males can establish exclusive territories within which only one male is dominant. Each dominant male defends his territory against incursions by other males and allows one or two females to cohabit with him (Crowcroft and Rowe, 1963; Crowcroft, 1955; Anderson and Hill, 1965). Reimer and Petras (1967) built a complex pen and observed that subordinate males sometimes shared the territory of a dominant male. They also found that groups of females established territories (which were defended against other females) and although male and female territories were usually the same, sometimes female territories overlapped two or more male territories. If a strange mouse is introduced into an established mouse colony in a pen, it is attacked and often killed (Reimer and Petras, 1967; Oakeshott, 1974). Males are normally more aggressive than females and fight more often. Males rarely attack females (usually only accidentally) but pregnant females near parturition or lactating females can be extremely aggressive towards males in the vicinity of the nest (Brown, 1953; Crowcroft and Rowe, 1963). In colonies established with only one male no fighting occurs until the first litter matures. The young males begin fighting when two to three months old (Crowcroft and Rowe, 1963; Brown, 1953) and one of these may eventually defeat the original male, to become dominant (Brown, 1953; Reimer and Petras, 1967).

Confined mouse populations usually grow rapidly, but eventually population growth slows and stops. Many density-dependent factors have been suggested to slow the growth of confined populations: reduced maternal care because of frequent disturbance to lactating females causing infant mortality (Brown, 1953; Southwick, 1955a,1955b; Reimer and Petras,1967; Lloyd and Christian, 1967; Christian, 1963); cannibalism (Southwick,1955a; Lloyd and Christian, 1967); reduction in the proportion of adult females becoming pregnant (Southwick, 1955a; Crowcroft and Rowe, 1958; Christian, 1963; Lloyd and Christian, 1969); increased mortality of weaned mice (Southwick, 1955a; Reimer and Petras, 1967); reduced food intake (Southwick, 1955a); increased proportion of anoestrous adult females (Crowcroft and Rowe, 1957,1958); decreased rate of implantation of embryos (Christian, 1963); and increased intrauterine mortality (foetal resorption) (Christian, 1963; Lloyd and Christian, 1969).

Socially dominant mice remain in breeding condition for longer than subordinates who either cease breeding or never start breeding when population densities plateau. (Lloyd and Christian, 1969). Crowcroft and Rowe (1958) showed experimentally that the high density of mice in a pen caused the cessation of breeding in mouse colonies which stopped growing in number. On all three occasions when mice were allowed to spread from small, overcrowded pens into much larger pens, successful reproduction occurred shortly after the reduction in density.

6.3 THE GENETICS OF WILD POPULATIONS OF HOUSE MICE

6.3.1 Amount of Genetic Variation

Following the demonstration by Lewontin and Hubby (1966) that 7 out of 18 electrophoretically detectable loci showed genetically controlled electrophoretic variation in *Drosophila pseudoobscura*, there have been

many papers giving estimates of the proportion of polymorphic loci (as detected by electrophoresis) in a range of organisms. Selander and Yang (1969) and Selander *et al.* (1969) caught house mice from three subspecies in a survey for electrophoretically polymorphic loci. In Californian mice, 12 of 40 loci were polymorphic; in Northern Denmark, 15 of 41 loci were polymorphic; and in Southern Denmark, 9 of 41 loci were polymorphic. On average, in each mouse population sampled, individuals were heterozygous at 8.5% of their loci (Selander, *et al.*,1969). Ruddle *et al.* (1969) obtained similar results with mice from North Carolina, Vermont and Alberta: 5 of 17 loci were polymorphic. In Ontario, Petras *et al.* (1969) found 5 of 8 loci had electrophoretically detectable polymorphisms, but since this survey was of loci already identified as polymorphic on wild or domestic stocks of mice, there was a bias towards polymorphic loci in this study.

Prior to the application of electrophoretic techniques to the search for polymorphisms, other genetically controlled polymorphisms were known to be present in wild populations of mice. The *Agouti* locus has two alleles, A^w and A^+ (see Chapter 8), and many populations contain both alleles (e.g. Dunn *et al.*,1960; Petras *et al.*,1969). Numerous *t*-alleles (usually denoted t^{wx}, where x refers to the xth *t*-allele extracted from wild populations) at the Brachy (*T*) locus have been recovered from wild populations. Fresh *t*-alleles from wild populations are detected by their effects on tail length when heterozygous with the dominant *T* allele (a mutant of laboratory origin known only in house mice, so *t*-alleles cannot be found in other species). The tails of *T/+* mice are shorter than normal, *T/t* mice are tail-less, *+/t* and *t/t* (if viable) mice have normal tails and *T/T* mice die *in utero*. Different *t*-alleles are either recessive lethals or cause male (but not female) sterility when homozygous. The *t*-alleles are maintained in wild populations because the segregation ratio

in heterozygous males is not 1*t*:1+, but is biassed heavily in favour of the *t*-allele. The proportion of *t*-alleles in the sperm of a heterozygous male is usually .90 to .99, the exact frequency depending upon the particular *t*-allele and the genetic background (Dunn *et al.*, 1960; Johnston and Brown, 1969).

Brown (1965) observed a wild mouse colony containing homozygous pink-eyed dilute mice (*pp*).

Quantitative traits, in which part of the variation between individuals is due to genetic differences, have been studied in wild mice. Bader and Lehmann (1965) measured the amount of phenotypic variation in the width of each of the three molar teeth of mice in different populations. The least variation between mice within a population was found in F_1 hybrids formed by crossing two inbred lines. The inbred lines were more variable than their hybrids, suggesting that more heterozygous mice are better buffered against environmental variation than more homozygous mice ("developmental homeostasis", Lerner 1953). The most variable populations were of wild mice, evidence for greater environmental and genetic variation between wild mice than domestic mice.

Another class of variants that have been intensively studied in mice are the discontinuous skeletal variants. Many skeletal characters can occur in either of two states (e.g. a fusion of two bones may or may not occur; a foramen may be present or absent) and each population has a certain frequency of each state. These skeletal variants (called epigenetic polymorphisms by Berry, 1963) were originally described in domestic mice, but Weber (1950) and Berry (1963,1964) showed that they occurred in wild mice.

Petras (1967c) surveyed mouse populations in Michigan for recessive alleles at five loci affecting coat colour. 82 wild mice were crossed to tester stocks that were homozygous for recessive alleles at these loci, but none of the wild mice were heterozygous at any of these loci. In

other crosses, a wild mouse was mated to a domestic mouse and then remated to seven of the progeny to check for recessive alleles carried by the wild parent. This series of crosses yielded one family with white spotting on dorsal and ventral regions, suggesting that recessive genes affecting coat colour are occasionally present in wild mouse populations.

6.3.2 Selection in Mouse Populations

The simplest case of selection in house mouse populations was the selective predation of pinkeyed dilute (*pp*) mice observed by Brown (1965). When this mouse population was protected from cats, the frequency of *pp* mice increased to 27/58 (46.6%). After cats were given access to these mice, the frequency of the pale and conspicuous *pp* mice declined to zero within three months, although many of the normal mice survived. The recessive allele (*p*) was not eliminated because when the cats were excluded after preying on the mice for nine months, more *pp* mice appeared in the next generation of juveniles.

Van Valen (1965) observed natural selection acting on the width of the first upper molar in wild mice from many populations in the British Isles. This character does not alter after a mouse has matured, consequently any changes in the mean and variance of the character in older age classes is evidence for differential mortality among mice with different molar widths. In most populations, Van Valen found that older mice have larger mean molar widths and are more variable than younger mice. Thus natural selection was favouring mice with more extreme (and especially larger) phenotypes (i.e. "destabilising" or "divergent" selection).

Some research has been done on the relative fitnesses of the genotypes at the *t*-locus. Dunn *et al.* (1958) suggest that heterozygotes for recessive lethal *t*-alleles may be fitter than the normal homozygote. For male sterile *t*-alleles, Dunn and Bennett (1966) reported that $t^{wx}t^{wx}$ embryos were less viable than their T/t^{wx} sibs and Johnston and Brown (1969)

estimated that the fitnesses of +/+ and t^{w2}/t^{w2} were 1.351 and 0.91 respectively, relative to $+/t^{w2}$.

Selander (1970) observed that samples of mice from small subpopulations showed an excess of heterozygotes relative to the number expected from Levene's (1949) formula for small samples,at the *Hbb* and *Es-3* loci. This excess of heterozygotes was regarded as evidence for selection (heterosis), but as was discussed in Chapter 3, this can also be explained by inbreeding in small subpopulations.

Berry and Murphy (1970) found seasonal changes in the genotype frequencies at two loci in the mice of Skokholm Island. At the *Hbb* locus, there was a highly significant excess of heterozygotes (Hbb^dHbb^s) in autumn and only a slight excess in spring. At the *Es-2* locus, there was a highly significant excess of heterozygotes ($Es\text{-}2^aEs\text{-}2^b$) in males (very slight excess in females) in spring and a barely significant excess in autumn. The same genotype frequency changes occurred at both loci in two successive years. Unfortunately, I have found both of these loci difficult to score reliably (Chapter 8): $Es\text{-}2^aEs\text{-}2^b$ heterozygotes are often difficult to distinguish from the homozygotes and environmental factors may alter the phenotypes, and Hbb^dHbb^s is difficult to distinguish from Hbb^dHbb^d in haemolysates that have been frozen and thawed (Selander, Young and Hunt, 1969; Wheeler and Selander, 1972). Berry and Murphy do not provide adequate details of how they scored the phenotypes at these loci, nor do they mention making crosses to confirm the genotypes of mice with different phenotypes. Such details and genetic studies would make their evidence for selection more convincing. The fact that none of the other major studies of these loci in wild mice (Petras, 1967a; Selander, Yang and Hunt, 1969; Wheeler and Selander,1972) found similar evidence of selection (apart from the *Hbb* data already discussed above) might not be relevant because the mice on Skokholm are living under very different conditions from the mice in the other studies.

Berry and Murphy (1970) found evidence for directional selection acting on the *Pep-C* locus when the frequency of the *Pep-C2* allele declined from 92% in spring, 1968, to 66% in autumn, 1969, and the frequency of *Pep-C3* increased from 6% to 26%.

Good evidence for selection in house mouse populations comes from studies of the hybrid zone between two subspecies of house mice (*M.m. domesticus* and *M.m. musculus*) in Denmark, which show that: (1) the zone has been stable for more than 20 years; (2) at only 4 of 17 polymorphic loci are there similar gene frequencies in both subspecies; (3) at 6 of 17 polymorphic loci, different alleles are fixed (or nearly so) in the two subspecies beyond the range of the hybrid zone; (4) the hybrid zone is only about 20 Km wide in the narrowest region (Selander, Hunt and Yang, 1969; Hunt and Selander, 1973). The fact that these two subspecies can remain genetically distinct despite contacts which are believed to have lasted several thousand years, strongly supports the view that the genes and gene combinations (within chromosomes) typical of each subspecies have reduced fitness when combined (in hybrids and backcrosses) with genes and chromosomes from the other subspecies. However, this type of study does not discriminate between selection acting on individual loci or on co-adapted gene complexes (Mayr, 1963).

6.3.3 Genetic Evidence for Population Subdivision

Berry (1963) measured the frequency of skeletal variants (epigenetic polymorphisms) in subpopulations of mice within three populations: 15 corn ricks on a farm in Hampshire; eleven localities in different parts of the British Isles; and nine other places throughout the world. He found that the rick subpopulations differed from each other and these differences were probably mainly genetic in origin. Whilst selection cannot be ignored as an explanation for the differences between rick subpopulations, the small number of founders of each subpopulation

suggests that random sampling drift is a likely explanation for these differences. The differences between subpopulations within the British Isles and over the world were highly significant, but were not as large as Berry expected by extrapolation from the differences between ricks on a farm and he thought that stabilising selection was acting upon the epigenetic polymorphisms over wide areas.

Berry (1964) used epigenetic polymorphisms to show that mice on two islands (Skokholm and May) were distinctively different from mice on the nearby mainland. There was no evidence of convergent evolution of the two island populations. Berry concluded that the peculiar features of the island mice stemmed from characteristics that were present or absent, by chance, in the founders of each population.

Petras (1967a) used electrophoretic variants as a method for detecting population subdivision in house mice. He used the methodology of Rasmussen (1964), who had previously used a blood group polymorphism in deer mice to show how a deficiency in heterozygotes relative to Hardy-Weinberg frequencies was evidence for population subdivision (providing no recessive null alleles are present). A source of bias exists in Petras' study: if there were seasonal or long term changes in gene frequency then pooling data from different seasons and years (as he did) could lead to an over-estimate of the amount of inbreeding due to population subdivision (Selander, 1970). Inbreeding coefficients (F_{IT}) estimated from genotype frequencies at three loci were consistent: 0.18, 0.13 and 0.15 for the *Es-2*,*Hbb* and *A* loci, respectively (Petras, 1967a,1967c). Petras compared these estimates of F_{IT} with theoretical values of F_{IT} for the "island" and "isolation by distance" models of population structure. The equations that he used were for monoecious, not dioecious, populations (c.f. Chapter 3), but the error due to this approximation is probably minor. By assuming that the average immigration rate into a mouse subpopulation was between 0.02 and 0.20,

Petras estimated that the effective size of the basic, panmictic subpopulation was between 6 and 80 mice. After a review of the ecological literature, he decided that these estimates of the panmictic subpopulation size in buildings and granaries were valid.

Selander (1970) summarised the research that he and his associates had done on population structure in house mouse populations in Texas. In his review of the literature on the ecology and behaviour of house mice he concluded that the effective breeding size of the average panmictic subpopulation of house mice was probably less than ten. Selander gave genetical evidence for population structuring within large barns and between barns on the same farm. When large chicken barns were split into equally sized sub-units, he found that gene frequencies were highly heterogeneous between sub-units. Evidently, in these chicken barns, there is a stable and dense (up to 7/sq. metre) mouse population that subdivides the habitat by the establishment of defended territories which prevent gene flow through each barn. Unfortunately, Selander does not substantiate his genetical analysis of these populations with adequate ecological data on the extent of individual territories, birth rates and mortality rates, dispersal rates and dispersal distances.

Gene frequencies in neighboring barns were highly significantly heterogeneous, suggesting that movement of mice between barns was not sufficient to eliminate genetic differences between barns. Selander interpreted this observation as evidence for fierce territoriality that prevented dispersing mice from entering densely inhabited areas. Because he did not estimate effective population sizes within barns and immigration rates between barns, it is not possible to refute the suggestion that the genetic differences between barns are due to selection.

Petras *et al.* (1969) collected data on 6 polymorphic loci from eight farms in Ontario. There were marked gene frequency differences between farms.

Studies of the hybrid zone in Denmark (Selander, Hunt and Yang, 1969; Hunt and Selander, 1973) suggest minimal amounts of gene flow between mouse subpopulations in neighboring farms. But here the situation is complicated by the action of selection, and until the magnitudes of the selective differences across the hybrid zone are known, it is not clear if the steepness of the hybrid zone is due to strong selection, weak gene flow or both strong selection and weak gene flow.

Population subdivision has been proposed by Lewontin and Dunn (1960) and Lewontin (1962,1968) as an explanation for the fact that observed frequencies of t-alleles in wild mouse populations are much lower than expected from simple mathematical models. The simplest mathematical models of t-allele in populations assume that all genotypes (except t^w/t^w) are equally fit and that there is no population subdivision (Bruck,1957; Dunn and Levene, 1961). Under these circumstances, lethal t-alleles should have gene frequencies about .35-.40 and male sterile t-alleles should be about .70. However, surveys of wild mouse populations yield much lower figures: Dunn *et al.* (1960) found 17.4% t-alleles in subpopulations that contained t-alleles (and many did not); Anderson (1964) found 12% t-alleles in neighboring barns in Alberta; and Petras (1967b) found 16% t-alleles in Michigan.

Recent data suggest that the relative fitnesses of the $+/+$, $+/t^w$ and t^w/t^w genotypes are not equal for male sterile t-alleles (Dunn and Bennett, 1966; Johnston and Brown, 1969) and these fitness differences may explain part (or all) of the discrepancy between simple theory and observation. Lewontin (1968) modified the mathematical theory of viable t-alleles to include viability differences between genotypes. The relative viabilities of t^w/t^w embryos observed by Dunn and Bennett (1966) were sufficiently low to halve the expected equilibrium gene frequencies (Lewontin,1968). Johnston and Brown (1969) found that their estimates of the segregation ratio in males and the relative viabilities of the three genotypes would

not allow a stable equilibrium gene frequency (the t^{w2} allele would be eliminated). Previously, Dunn and Levene (1961), had studied the t^{w2} allele in a confined colony of wild mice for seven years. The observed equilibrium gene frequency of t^{w2} was .37, approximately half the value predicted by mathematical theory that ignored fitness differences. In this case population subdivision could not be an explanation for the discrepancy between theory and observation: the average fitness of the t^{w2} allele must have been less than the average fitness of the + allele (ignoring the segregation ratio in males).

Levin *et al.* (1969) simulated a set of semi-isolated, finite subpopulations containing lethal t-alleles with 95% transmission from male heterozygotes. Random sampling drift within subpopulations could significantly reduce mean t-allele frequencies only when the effective subpopulation size was less than 8 when immigration rates were 1%, or less than 4 when immigration rates were 3%. Because effective subpopulation sizes and immigration rates in the wild are likely to exceed the values at which equilibrium gene frequencies are reduced by population subdivision, population subdivision is not an adequate explanation for the low frequency of t-alleles in wild populations.

The dispersal of t^{w11} through the mouse population on Great Gull Island (New York) after release at the eastern end in 1957 has been quoted as evidence for population subdivision that prevented the free flow of genes through a mouse population (Anderson *et al.*,1964). However, the data, as presented in a series of maps showing the genotypes and capture locations of mice tested for t-alleles in 1959, 1960, 1961 and 1962, do not support the assertion that the gene "... seems to be spreading slowly, since in the western part only two heterozygotes were found in 1961 and one other in 1962" (Anderson *et al.*,1964). What this statement ignores is the fact that on the eastern end of the island, there were only three heterozygotes in 1961 and one in 1962, that is, there were no differences

between the two ends of the island in 1961 and 1962. Furthermore, in 1961 the most westerly mouse caught was a t^{w11} heterozygote, an observation that does not support the contention that the gene was spreading slowly from the eastern end. Trapping was inadequate in 1959 and 1960 to do more than show that t^{w11} was established on the island.

Anderson (1964) and Petras (1967b) have collected data on the frequency of t-alleles within buildings on farms in Alberta and Michigan, respectively. Because of the difficulties involved in testing each mouse for t-alleles, sample sizes are too small for χ^2-tests for heterogeneity between buildings within farms. When the data from the larger farms in Alberta (Anderson, 1964) are pooled within farms, there is heterogeneity between farms.

6.4 DISCUSSION

This review of the literature on house mouse ecological genetics shows that the mouse is an excellent mammal with which to study subdivided populations. The species is common, occupies a wide range of habitats, has many genetic loci that are known to be polymorphic in wild populations, has had many aspects of its ecology investigated and many features of mouse behaviour are known.

Most genetical studies of natural populations have shown that the populations are subdivided, but many aspects of population subdivision are not fully understood. The efficiency with which the territoriality observed in confined colonies can restrict gene flow in wild populations is unknown. Although much is known about the home ranges and daily ranges of resident wild mice, the distance moved by dispersing mice is unknown. The ecological studies showed that mice have different ecological parameters (population density, reproductive rates, seasonal breeding, etc.) in different environments, but genetic studies have not looked at the genetic consequences of these different ecologies. Most of the genetic studies have been "snap-shots" (each subpopulation studied once) rather

than "movies" so there is a need for mark-release studies that regularly census mouse populations for several generations. Such studies could give valuable information on effective population sizes and immigration rates. Genetic differences have been detected between neighboring sub-populations; do these remain constant or do they vary as environmental changes alter the mouse population densities and dispersal rates?

CHAPTER 7

METHODS

7.1 STATISTICAL METHODS

Statistical methods were taken from Snedecor (1966). Yate's correction was applied in Chi-square tests, where appropriate. Computer programs were written in FORTRAN and run on the University of Adelaide CDC 6400 computer.

7.2 CAPTURE AND MAINTENANCE OF MICE

Wild mice were caught by hand (when very numerous) or in live traps ("Longworth" or "Sherman"). Matings were made up in ordinary mouse cages of tin or plastic, containing sawdust and shredded paper. The mice were mainly fed on "Mouse Pellets" supplemented with a weekly ration of birdseed. Female wild-caught mice rarely breed in captivity but most wild-caught males bred when mated with females from domestic laboratory stocks.

7.3 COLLECTION OF BLOOD SAMPLES

Each mouse was first anaesthetised with ether and a blood sample obtained either by cutting a jugular vein and collecting the blood in a heparinised tube or by puncturing the suborbital sinus with a needle and withdrawing blood into a heparinised syringe. The whole blood was spun in a centrifuge at 1-2,000g for 20 minutes and the plasma pipetted off, to be frozen at -30^{o}C. The erythrocytes were washed three times with 0.87% NaCl to remove all traces of plasma and the packed erythrocytes were frozen at -30^{o}C. Later the erythrocytes were thawed, mixed with an equal volume of 0.001-0.005M $CaCl_2$ and shaken to complete lysis. After centrifugation (at least 2,000g for 30 minutes) haemolysates were ready for electrophoresis.

7.4 ELECTROPHORESIS

Starch gel electrophoresis was carried out in gel trays that were 30 cm x 15.5 cm x 6mm with "Connaught" starch. Slot formers, projecting down from plastic lids that covered the gel trays, were used to make one or two rows of slots across each gel, with 8, 12, 16 or 24 slots in each row. Thin slots (< 0.5 mm thick) were used for plasma samples and thicker slots (1 mm thick) were used for haemolysates. Two electrophoresis buffer systems were used: Gahne's (1966) Tris-citrate/Lithium borate (pH 8.5) was best for plasma, and Smithies' (1959) Tris-EDTA-Borate (pH 8.6) for haemolysates.

The staining methods used for detecting enzymes were usually taken from Shaw and Prasad (1970). In particular, proteins were stained with Amido Black (a 1% solution in 5:5:1 of water: methanol: acetic acid, respectively) and esterases were detected using a mixture of α and β-Naphthyl-acetate in 0.05M Tris-Maleate (pH 6.5) buffer with Fast Garnet GBC as dye-coupler.

Careful treatment was required to successfully score each of the three loci detected by electrophoresis of haemolysates. The *Es-3* locus (see Chapter 8) formed multiple subsidiary bands when too much heparin was present (Templeton, 1969) and these extra bands made scoring of the phenotypes very difficult. Consequently, only small amounts of heparin were used in blood samples. The addition of $CaCl_2$ to haemolysates and to the esterase staining solution helped to overcome problems caused by heparin and improved the staining of esterases. Another problem with scoring for *Es-3* phenotypes occurred when plasma contaminated a haemolysate. In such a case, the very strongly staining esterase band (*Es-1*) in plasma showed up on starch gels in the region of the $Es\text{-}3^d$ allele and could be confused with it. To avoid this problem, erythrocytes were washed thoroughly and any blood clots removed before lysing the cells.

The heterozygotes at the *Hbb* locus (Hbb^dHbb^s) could not be reliably distinguished from Hbb^dHbb^d homozygotes until Dr. M. Petras suggested that the haemolysates be treated with maleate before electrophoresis. The following procedure worked successfully: 40µl of haemolysate were mixed with 10µl of 1M Tris-Maleate (pH 6.5) buffer and left overnight at 5^oC before electrophoresis. After this treatment, the Hbb^sHbb^s samples had a single slow protein band, Hbb^dHbb^d samples had a fast band and a very slow minor band, while Hbb^dHbb^s samples had both fast and slow bands and often showed the very slow minor band, particularly in old samples.

The *Erp-1* locus was best resolved after haemolysates were frozen and thawed several times. Further details about the methods used to score the *Erp-1* phenotypes are given in Appendix 1.

CHAPTER 8

THE INHERITANCE OF POLYMORPHISMS AT TEN LOCI FROM NATURAL POPULATIONS OF THE HOUSE MOUSE IN SOUTH AUSTRALIA.

8.1 INTRODUCTION

Although allelic variation has been recorded at over 500 gene loci in the house mouse (Searle, 1974), not many of these loci are found to be polymorphic in wild populations. Only one morphological polymorphism (involving the *agouti* locus) can readily be scored in South Australian mice. In order to score polymorphisms controlled by a number of separate loci, the proteins and esterases of the plasma and of the erythrocytic haemolysates were studied electrophoretically. All but two of the polymorphisms considered herein have been described previously and analysed genetically. The new polymorphisms discovered in the course of this study are an erythrocytic protein variant (*Erp-1*) and a coatcolour variant (*Ga*).

To compare the polymorphisms observed in South Australian mouse populations with those found in mouse populations in other countries, two checks must be made. Firstly, for each locus, the phenotypes must be identified by comparison with the phenotypes of standard inbred strains of mice. Secondly, the inheritance of the different phenotypes (as published) must be confirmed in the South Australian mice.

To identify the genes controlling the different phenotypes at each locus observed in wild mice, substrains of the following inbred lines were obtained: Balb/c, CBA, C3H, C57BL, DBA and SWR.

The genetical basis of the polymorphisms in wild mice were studied by mating wild mice to Genetics Department laboratory mice (usually laboratory ♀ x wild ♂) and scoring the progeny. Subsequently, crosses were made using F_1 and later generations and there were also a few successful matings of wild x wild mice.

TABLE 8.1 Results of crosses involving white (A^w.) and grey (A^+A^+) bellied mice.

♀ x ♂	Mating numbers	Progeny phenotypes		χ^2_1
		white belly	grey belly	
(1) aa x A^w.	1	8	1	
or A^w.x aa	2	5	8	
	4	8	6	
	5	5	3	
	Total	26	18	1.11(1:1)
	2 matings	36	0	
(2) A^+A^+ x A^w.	28	12	6	
or A^w. x A^+A^+	30	13	6	
	62	1	4	
	63	5	6	
	89	2	2	
	Total	33	24	1.12(1:1)
	4 matings	52	0	
(3) A^w. x A^w.	4 matings	27	0	
(4) A^+A^+ x A^+A^+	8 matings	0	98	
(5) A^+A^+ x A^wa	18	12	11	0.00(1:1)
(6) A^wA^w x A^wA^+	14	31	0	
(7) A^wA^+ x A^+A^+ or A^+A^+ x A^wA^+	3 matings	31	32	0.00(1:1)
(8) A^wA^+ x A^wA^+	2 matings	39	10	0.33(3:1)

8.2 RESULTS

8.2.1 *Agouti* locus (*A*). Chromosome 2.

Many house mouse populations around the world are polymorphic for white v. grey belly colour (Schwarz and Schwarz, 1943; Petras *et al*,1969). Although the two phenotypes vary considerably (especially in the amount of yellow pigment present) in my experience they can usually be distinguished from one another with little difficulty. Two alleles at the *agouti* locus are involved: white belly is due to the presence of the A^w allele while homozygosity for the A^+ allele results in grey belly (Gruneberg, 1952). A third allele, *a*, non-agouti, (*aa* mice are all of one colour, usually all black if combined with *B* at the brown locus) was used in some crosses because it is recessive to both A^+ and A^w, hence mice that are A^wA^+ should have white and grey bellied progeny when crossed to an *aa* mouse. In some crosses, the albino (*c*) gene was segregating and as albino (*cc*) mice cannot be scored for *agouti* and *c* is on chromosome 7 so that *c* should segregate independently of *A*, these albino mice have been left out of the data. Because the genotype of wild caught white bellied mice is not known *a priori*, these mice are denoted A^w, in Table 8.1, which presents the relevant results of crosses involving white and grey bellied mice.

Crosses (1) and (2) demonstrated that some white bellied wild caught mice were carrying the A^+ gene and crosses (5) to (8) show that A^wA^+ mice are white bellied, confirming that A^w is dominant to A^+. The segregation of white and grey bellied mice in crosses (1) and (2) shows a deficiency of grey bellied mice. If the results of matings segregating for grey belly in crosses (1) and (2) are pooled, the χ^2_1 for a 1:1 ratio is 2.55 (Yates correction). It should be noted that the expectation of a 1:1 ratio in matings segregating for grey belly is not exactly correct because the sample of matings chosen for analysis is biassed since matings involving A^wA^+ mice that do not yield grey progeny are ignored. Hence

TABLE 8.2 Results of crosses involving diffuse (Hbb^d) and sharp (Hbb^s) haemoglobins.

(a) When only diffuse and sharp phenotypes could be distinguished.

Cross ♀ x ♂	Mating numbers	Progeny phenotype		χ^2_1
		sharp	diffuse	
(1) $Hbb^d.$ x $Hbb^d.$	5 matings	0	67	
	11	1	9	
(2) $Hbb^d.$ x Hbb^sHbb^s	3	0	21	
	8	0	2	
	15	2	4	
(3) Hbb^sHbb^s x Hbb^sHbb^s	2 matings	36	0	
(4) Hbb^sHbb^d x Hbb^sHbb^d	2 matings	7	27	0.16(3:1)
(5) Hbb^sHbb^d x Hbb^sHbb^s	25	5	4	

(b) When all three phenotypes could be distinguished after treatment with tris-maleate buffer.

		Progeny phenotypes*		
		dd	*ds*	*ss*
(6) Hbb^sHbb^s x Hbb^sHbb^s	2 matings	0	0	36
(7) Hbb^dHbb^d x Hbb^dHbb^d	13	5	0	0
(8) Hbb^dHbb^d x Hbb^sHbb^s	3	0	11	0
(9) Hbb^sHbb^d x Hbb^sHbb^d	2 matings	3	5	3
(10) Hbb^dHbb^d x Hbb^sHbb^d	4	1	4	0

* *dd* = phenotype of Hbb^dHbb^d, etc.

more grey bellied than white bellied progeny should be expected and the real discrepancy between observation and expectation is greater than for the approximate 1:1 expectation. However, the results of crosses (7) and (8) suggest that the segregation of grey belly is normal and that the earlier deficiency is due to chance.

8.2.2 Haemoglobin beta chain (*Hbb*). Chromosome 7.

Electrophoretic variation in mouse haemoglobins was first described by Ranney and Glueckson-Waelsch (1955) using filter paper electrophoresis. On starch gels, Popp and St Armand (1960) found that the "diffuse" phenotype was dominant to the "single" phenotype. Shortly afterwards, Hutton *et al.* (1962) and Popp (1962) showed that this variation was due to amino acid differences in the haemoglobin beta chains, and the locus controlling this variation was named *Hbb*. The diffuse phenotype was produced by the presence of the Hbb^d allele, whilst the single or "sharp" phenotype was found in homozygotes for the Hbb^s allele. A third allele, Hbb^p, has been found in one inbred line (Morton 1962), but has not been recorded in any wild populations.

Later workers have found that if fresh haemolysates are used, the Hbb^dHbb^s heterozygote can be distinguished electrophoretically from the diffuse homozygote (Petras, 1967a; Selander *et al*, 1969; Berry and Murphy, 1970). However, for a long time I was unable to emulate their success. The buffer system of Petras and Martin (1969) did not discriminate between heterozygous and homozygous diffuse samples with my material and the biochemical method of Hutton (1969) was too complicated for routine work. In the results presented in Table 8.2, part (a), mice with the diffuse phenotype, but unknown genotype, are denoted by Hbb^d.. These results show that under the given conditions Hbb^d is dominant to Hbb^s. After the above work was completed, Dr. M. Petras kindly suggested that maleate should be added to haemolysates prior to electrophoresis

TABLE 8.3 Results of crosses involving the *Es-1* locus.

Cross ♀ x ♂	Mating Number	Progeny genotypes $Es\text{-}1^{b}Es\text{-}1^{b}$	$Es\text{-}1^{a}Es\text{-}1^{b}$	$Es\text{-}1^{a}Es\text{-}1^{a}$	χ^2_2
(1) $Es\text{-}1^{b}Es\text{-}1^{b}$ x $Es\text{-}1^{b}Es\text{-}1^{b}$	9 matings	99	0	0	
(2) $Es\text{-}1^{b}Es\text{-}1^{b}$ x $Es\text{-}1^{a}Es\text{-}1^{a}$	61	0	15	0	
(3) $Es\text{-}1^{a}Es\text{-}1^{b}$ x $Es\text{-}1^{b}Es\text{-}1^{b}$	89	2	2	0	
(4) $Es\text{-}1^{a}Es\text{-}1^{b}$ x $Es\text{-}1^{a}Es\text{-}1^{b}$	4 matings	19	40	18	.06(1:2:1)

and this simple procedure made it possible to classify *Hbb*^{d}. mice as heterozygotes or homozygotes. The results from treating some haemolysates with tris-maleate buffer as outlined in Chapter 7 are given in Table 8.2, Part (b). The three genotypes are readily scored by this method.

8.2.3 Esterase-1 locus (*Es-1*). Chromosome 8.

Electrophoretic variation of the major plasma esterase in the mouse was detected using starch-gel electrophoresis by Popp and Popp (1962). Petras and Biddle (1967) showed that *Es-1* is closely linked to *Es-2* and *Es-5*. Selander, *et al* (1969) reported the existence of two very rare alleles in wild populations, *Es-1*^{d} (with a mobility different from *Es-1*^{a} or *Es-1*^{b}), and *Es-1*^{c} which is a null allele. The results of crosses involving the *Es-1* locus are given in table 8.3, confirming the simple inheritance of *Es-1*^{a} and *Es-1*^{b}.

8.2.4 Esterase-3 (*Es-3*). Chromosome 11.

When serum-free mouse haemolysates are run on alkaline starch gels one or more bands of esterase activity can be found ahead of the haemoglobin zone. These bands are controlled by the *Es-3* locus, which was first described from kidney extracts by Ruddle and Roderick (1965) and later in haemolysates by Pelzer (1965) who, however, called the locus *Ee-2*. Popp (1966) demonstrated that *Es-3* and *Ee-2* were the same locus and reported three alleles in a survey of 22 inbred lines. *Es-3*^{a} is a silent (or null) allele which when homozygous leads to an almost complete absence of esterase activity, *Es-3*^{b} produces an esterase band of intermediate mobility and *Es-3*^{c} produces a slow moving band in alkaline gels. The heterozygotes, *Es-3*^{a}*Es-3*^{b} and *Es-3*^{a}*Es-3*^{c}, have only a single band of esterase activity while *Es-3*^{b}*Es-3*^{c} has two bands. A fourth allele, *Es-3*^{d}, which is often found in low frequencies in wild

TABLE 8.4 Results of crosses involving the *Es-3* locus.

Cross ♀ x ♂	Mating number	Progeny genotypes*					χ^2_1
		cc	*cb*	*bb*	*cd*	*dd*	
(1) $Es\text{-}3^{c}Es\text{-}3^{c}$ x $Es\text{-}3^{c}Es\text{-}3^{c}$	4 matings	39	0	0	0	0	
(2) $Es\text{-}3^{c}Es\text{-}3^{c}$ x $Es\text{-}3^{b}Es\text{-}3^{c}$ or $Es\text{-}3^{b}Es\text{-}3^{c}$ x $Es\text{-}3^{c}Es\text{-}3^{c}$	7 matings	41	49	0	0	0	.54(1:1)
(3) $Es\text{-}3^{b}Es\text{-}3^{b}$ x $Es\text{-}3^{b}Es\text{-}3^{b}$	11	0	0	10	0	0	
(4) $Es\text{-}3^{c}Es\text{-}3^{d}$ x $Es\text{-}3^{c}Es\text{-}3^{d}$	13	5	0	0	6	2	
(5) $Es\text{-}3^{d}Es\text{-}3^{d}$ x $Es\text{-}3^{d}Es\text{-}3^{d}$	14	0	0	0	0	20	

* Progeny genotypes: *cc* = $Es\text{-}3^{c}Es\text{-}3^{c}$, etc.

TABLE 8.5 Results of crosses involving the *Es-5* locus.

Cross ♀ x ♂	Mating number	Progeny phenotypes		χ^2_1
		Es-5^b.	*Es-5*^{a}*Es-5*^{a}	
(1) *Es-5*^{b}. x *Es-5*^{b}.	3	6	11	** 12.3(3:1)
	7 matings	72	0	
(2) *Es-5*^{b}. x *Es-5*^{a}*Es-5*^{a}	4	8	3	
or *Es-5*^{a}*Es-5*^{a} x *Es-5*^{b}.	7	4	2	
	14	16	8	
	Total	28	13	* 4.8(1:1)
	2 matings	43	0	
(3) *Es-5*^{a}*Es-5*^{a} x *Es-5*^{a}*Es-5*^{b}	3 matings	24	21	.09(1:1)
(4) *Es-5*^{b}. x *Es-5*^{a}*Es-5*^{b}	25	9	0	
	15	5	1	
(5) *Es-5*^{a}*Es-5*^{a} x *Es-5*^{a}*Es-5*^{a}	11	0	10	

* Probability < 5%

** Probability < 1%

mouse populations in North America (Selander, *et al*, 1969; Martin and Petras, 1971), produces a faster band than *Es-3*^{b}. Selander, *et al* (1969) claim to have found another null allele, *Es-3*^{e}, in two wild mice but they do not mention how it differs from *Es-3*^{a} (which has otherwise not been found in wild mice). They also mention the presence of a *Es-3*^{f} allele in Denmark, but provide no details.

In South Australia, alleles thought to be *Es-3*^{b}, *Es-3*^{c} and *Es-3*^{d} have been observed in wild mice. Only *Es-3*^{c} has been definitely identified by scoring the haemolysates of CBA, C3H and DBA mice. Comparison of my gels with published diagrams (Selander, *et al*, 1969; Martin and Petras, 1971) indicates that the other two alleles are correctly identified.

The results of crosses made to confirm the inheritance of the *Es-3* alleles are given in Table 8.4. The three alleles are codominant.

8.2.5 Esterase-5 (*Es-5*). Chromosome 8.

The *Es-5* locus, first described by Petras and Biddle (1967) controls the presence of a plasma esterase band on gels. There are two alleles, *Es-5*^{b} which produces a single esterase band and *Es-5*^{a} which is a recessive null allele. Because of the variable presence or absence of other plasma esterase bands in mice from different populations, it is sometimes difficult to score this locus in wild mice. Plasma from C57BL mice was used as a standard when scoring for the *Es-5*^{b} esterase band.

The results of matings scored for *Es-5* phenotypes are presented in Table 8.5, where mice expressing the *Es-5*^{b} band, but of unknown genotype, are denoted by *Es-5*^{b}.. The result of mating (3) in cross (1) is clearly aberrant. The progeny suggest that the cross was actually *Es-5*^{a}*Es-5*^{a} x *Es-5*^{a}*Es-5*^{b}, which implies that one of the parents was mis-scored as *Es-5*^{b}.. Also, in cross (2), the ratio of *Es-5*^{b}. to *Es-5*^{a}*Es-5*^{a} in matings segregating for both phenotypes suggests that some *Es-5*^{a} homozygotes were

TABLE 8.6 Results of crosses of *Pre* phenotypes.

Cross ♀ x ♂	Mating number	Progeny phenotypes		χ^2_1
		$Pre^{o}Pre^{o}$	$Pre^{a}.$	
(1) $Pre^{o}Pre^{o}$ x $Pre^{a}.$	6	9	21	4.03(1:1)
	5 matings	28	21	0.74(1:1)
(2) $Pre^{o}Pre^{o}$ x $Pre^{o}Pre^{a}$	3 matings	28	15	3.35(1:1)
(3) $Pre^{a}.$ x $Pre^{a}.$	2 matings	0	31	
(4) $Pre^{o}Pre^{o}$ x $Pre^{o}Pre^{o}$	2 matings	22	0	
(5) $Pre^{o}Pre^{a}$ x $Pre^{o}Pre^{a}$	25	2	7	

TABLE 8.7 Segregation of *Pre* phenotypes in males and females in pooled data from crosses (1) and (2), excluding mating 6.

Sex	Phenotype		χ^2_1 (1:1 ratio)
	$Pre^{o}Pre^{o}$	$Pre^{a}.$	
♀	36	7	18.2**
♂	20	29	1.3
Total	56	36	4.0*

χ^2_1 for homogeneity = 15.9**

* Probability < 5%

** Probability < 1%

mis-scored as *Es-5*^{b}.. However, this bias is negligible in cross (3) where the excess of *Es-5*^{b}. phenotypes is insignificant. The data from cross (5) show full ascertainment of *Es-5*^{a}*Es-5*^{a} phenotypes but the sample size is rather small in this case. As a test eight mice were bled twice and their *Es-5* phenotype determined each time: 6 were clearly scored as *Es-5*^{b}. both times, one was clearly *Es-5*^{a}*Es-5*^{a} both times and one was clearly *Es-5*^{b}. the first time, but was difficult to score the second time and would probably have been scored as *Es-5*^{a}*Es-5*^{a}. It thus appears that some *Es-5*^{a}*Es-5*^{a} mice can be mis-scored as *Es-5*^{b}..

8.2.6 Pre-albumin (*Pre*)

Genetically controlled variation in the mouse plasma pre-albumin region was described by Shreffler (1964) who found two alleles in a survey of laboratory stocks. The *Pre*^{a} allele controls the production of a protein band in the pre-albumin region which is lacking in mice homozygous for the allele *Pre*^{o}. In his crosses between inbred lines Shreffler was able to distinguish *Pre*^{a}*Pre*^{o} and *Pre*^{a}*Pre*^{a} mice by the intensity of protein staining. However, in wild mice in South Australia, no clear-cut differentiation of *Pre*^{a}. phenotypes into two classes was possible and so *Pre*^{a} was scored as dominant to *Pre*^{o}. Shreffler found that there was a lower concentration of pre-albumin protein in females than in males and this made scoring of phenotypes in females difficult. His data show that a proportion of *Pre*^{a}*Pre*^{o} females were mis-scored as *Pre*^{o}*Pre*^{o}.

This locus is very difficult to score as electrophoretic and staining conditions must be nearly perfect before repeatable results can be obtained. Some samples had to be re-run two or three times before they could be scored confidently.

The results of crosses involving the *Pre* locus are given in Table 8.6. In cross (1), the results of mating (6) are significantly different from

TABLE 8.8 Population samples classified for *Pre* phenotype and sex.

Sample	Sex	$Pre^{o}Pre^{o}$	Pre^{a}.	χ_1^2 for homogeneity
Wardang Island	♀	23	11	0.62
	♂	64	46	
Price	♀	13	18	0.09
	♂	11	11	
South Kilkerran	♀	20	17	0.00
	♂	15	11	
Alford	♀	26	11	2.33
	♂	20	2	
Minlaton	♀	15	8	0.10
	♂	7	6	
Port Vincent	♀	12	22	1.31
	♂	12	10	
Turretfield, 1970	♀	22	16	0.08
	♂	10	5	
O'Halloran Hill	♀	21	13	0.00
	♂	31	17	
Turretfield, all mice July 1971 to May 1972	♀	150	111	10.19 Probability <1%
	♂	113	149	

those expected if the female were $Pre^{o}Pre^{o}$ and the male $Pre^{a}Pre^{o}$. However, mating (6) fits a 1:3 ratio ($\chi^2_1 = 0.18$) suggesting that the female was actually $Pre^{a}Pre^{o}$ and had been mis-scored. Table 8.7 shows that the observed segregation is normal in males, but highly significantly deviant in females, presumably because $Pre^{a}Pre^{o}$ females are frequently misclassified as $Pre^{o}Pre^{o}$ as in Schreffler (1964).

In view of the many differences that can be detected between various stocks of mice, it is reasonable to ask whether or not the sex difference observed in mice that are partly of laboratory origin and living in laboratory conditions is also found in wild mice. To check this point, mice from many populations have been classified by sex and *Pre* phenotypes and the data then tested for homogeneity in 2x2 tables. The results are given in Table 8.8. In the first eight populations no statistically significant deviation from homogeneity was observed and in fact there is a tendency to obtain more female Pre^{a}. phenotypes than expected, but in the last sample a highly significant deficiency of female Pre^{a}. phenotypes was observed.

No satisfactory conclusion can be drawn from these results. The data from laboratory crosses and the extensive field data from Turretfield conflict with the data from eight other populations. No simple explanation (such as variation in scoring techniques, differences in the ages of the plasma samples or a bias on my part as I scored the gels) can account for these differing results. Nevertheless, it is clear that phenotype frequencies in females must be treated very cautiously as there is a great risk of misclassification.

8.2.7 Transferrin (*Trf*). Chromosome 9.

Two alleles are known at the mouse transferrin locus. Trf^{b} (Cohen and Schreffler, 1961) occurs in most inbred lines and in all wild mouse

populations reported in the literature (*e.g.* Petras *et al*, 1969; Berry and Murphy, 1970) and Trf^a is found only in the CBA inbred line. A modified form of the Trf^b homozygote was reported by Ashton and Braden (1961) which can be confused with the Trf^aTrf^b heterozygote. Shreffler (1963) commented that young mice between one and eight weeks of age often have a prominent subsidiary *Trf* band that can cause Trf^bTrf^b mice to be mistaken for Trf^aTrf^b. This can also occur during disease attacks or other stresses.

Some wild caught South Australian mice had phenotypes resembling Trf^aTrf^b (the majority were clearly Trf^b homozygotes), but when two of them were bred in the laboratory, this phenotype did not appear in their progeny:

(a) mating 61: ♀ "*ab*" * ♂ *bb* → 16*bb*

(b) mating 88: ♀ *bb* * ♂ "*ab*" → 10*bb*

where "*ab*" denotes the supposed Trf^aTrf^b phenotype and *bb* = Trf^bTrf^b.

Consequently, it is assumed that many, if not all, of the apparent Trf^aTrf^b heterozygotes found in wild mice are in fact modified Trf^b homozygotes.

Berry and Murphy (1970) treated the modified Trf^b homozygote phenotype, or one very similar to it, as though it were genetically controlled, but in view of the negative results from the two matings given above, this is not a justifiable procedure with wild caught mice in South Australia.

8.2.8 Esterase-2 (*Es-2*). Chromosome 8.

Petras (1963) denoted the locus controlling the staining intensity of the most anodal plasma esterase band in alkaline starch gels by *Es-2*. He described a null allele ($Es\text{-}2^a$) and an allele associated with the presence of this esterase band ($Es\text{-}2^b$). These two alleles were codominant.

TABLE 8.9 Population samples scored for *Es-2* phenotypes in plasma.

Population Sample	Phenotypes (and possible genotypes)				F_{IS} after pooling $Es\text{-}2^c$ with $Es\text{-}2^b$
	no activity ($Es\text{-}2^a Es\text{-}2^a$)	Light staining ($Es\text{-}2^a Es\text{-}2^b$ $Es\text{-}2^a Es\text{-}2^c$)	Heavy staining ($Es\text{-}2^b Es\text{-}2^b$)	2 Light staining bands ($Es\text{-}2^b Es\text{-}2^c$)	
Turretfield,1970	82	54	4	1	$-.124 \pm .072$ **
Price	46	11	0	0	$-.107 \pm .032$
South Kilkerran	17	36	12	0	$-.114 \pm .123$
Alford	46	17	1	0	$-.051 \pm .111$
Minlaton	41	17	1	0	$-.067 \pm .112$
Port Vincent	45	15	3	0	$.143 \pm .146$
Port Lincoln	18	10	0	1	$-.234 \pm .069$
Wardang Island	40	83	21	0	$-.173^{*} \pm .081$
O'Halloran Hill	12	45	25	1	$-.136 \pm .108$

* Probability < 5%

** SE after Rasmussen (1964)

Ruddle *et al* (1969) described an electrophoretic mobility variant (produced by a third allele, $Es\text{-}2^c$) in which the esterase band moved more slowly in alkaline gels than the band produced by $Es\text{-}2^b$. All three alleles have been found in wild mouse populations; Selander *et al* (1969) described an additional four electrophoretic mobility variants (produced by another four alleles, presumably) in wild mice.

Samples of wild mice from South Australia showed that the $Es\text{-}2^a$, $Es\text{-}2^b$ and $Es\text{-}2^c$ alleles were present, although $Es\text{-}2^c$ was rare (Table 8.9). However, scoring of the three genotypes, $Es\text{-}2^aEs\text{-}2^a$, $Es\text{-}2^aEs\text{-}2^b$ and $Es\text{-}2^bEs\text{-}2^b$ was difficult because nearly perfect electrophoretic and staining conditions were essential to obtain repeatable results. It was possible to distinguish three phenotypes: (a) almost no esterase activity; (b) light staining; (c) heavy staining of the plasma *Es-2* esterase band. But there was some overlap between these classes and for some plasma samples, the final phenotypic designations were somewhat arbitrary. Nevertheless, population samples usually fitted Hardy-Weinberg frequencies (Table 8.9).

Crosses set up for genetical study of this system were shown to be useless by two observations. Firstly, the inbred lines (Balb/c, C57BL, CBA, DBA, SWR) used as standards had virtually no *Es-2* esterase activity detectable in their plasma, when all were supposed to have full $Es\text{-}2^bEs\text{-}2^b$ activity. Secondly, when three wild mice with intermediate activity and one with intense staining were brought back to the laboratory, all lost their *Es-2* plasma esterase activity after living under laboratory conditions for several months. Also, other mice living in the laboratory colony gave various phenotypes when scored on different occasions.

These observations suggested that the *Es-2* genotype was not correctly expressed as a phenotype in plasma samples from mice kept in the laboratory colony. Some C57BL, CBA and SWR kidney extracts were stained for esterases and strong *Es-2* esterase activity was observed

in each case. Ruddle *et al* (1969) found that their mice had the same phenotype in plasma and kidney, but this was not the case with mine. As might have been expected from these observations, the crosses set up to check the inheritance of *Es-2* phenotypes resulted in progeny phenotypes that were incompatible with the parental phenotypes according to the model presented by Petras (1963) of two codominant alleles at one locus and no environmental modifications of the phenotype .

The factor (or factors) in the environment of my mouse colony which caused the instability of the *Es-2* phenotypes in plasma has not been identified. Since I used the same inbred lines as Petras (1963) and Ruddle *et al* (1969) the differences in phenotypes is unlikely to be due to genetic differences in the mice, but must presumably be environmental and something unique to my colony. The food, water and bedding used in mouse cages are probably different in North America and South Australia and it is possible that one or more of these items could play a role in the expression of the *Es-2* locus in plasma. Nothing is known about the stability of the *Es-2* phenotypes in wild mice in their normal habitat because none were bled twice whilst living under wild conditions.

8.2.9 Erythrocytic protein-1 locus (*Erp-1*). Chromosome 8.

This locus which does not appear to have been described before, controls electrophoretic variation in an erythrocytic protein. The genetical data and analysis are presented in Appendix 1.

8.2.10 Greying with age (*Ga*)

This is another genetic variant which, it seems, has not been described previously. The *Ga* gene causes mature mice to become grey because of a progressive loss of pigmented hairs as they age. The genetical data are given in Appendix 2.

8.3 DISCUSSION

The genetical data collected for the 10 loci described above show that the *A*, *Hbb*, *Es-1*, *Es-3* and *Erp-1* loci have phenotypes that are reliably produced by certain genotypes. But the *Es-2*, *Es-5*, *Pre*, *Trf* and *Ga* loci have phenotypes that are affected by various environmental factors so that the genotype is not always reliably expressed in the phenotype. Consequently, the data collected from wild populations on the last five loci cannot be used for any genetic analysis of the mouse populations of South Australia, simply because the reliability of such data is questionable. Admittedly, for the *Es-2* and *Pre* loci, most of the evidence concerning the unreliability of the phenotypes comes from laboratory crosses rather than wild populations and so the phenotypes at these loci may not be altered by non-genetic factors in wild mice. But, if these data were to be used, doubt would always remain about the validity of any analyses based on them. For this reason, the extensive data collected on the *Es-2*, *Es-5* and *Pre* loci in particular, must be set aside and the genetic analyses presented subsequently in this thesis will use only the data on the *A*, *Hbb*, *Es-1*, *Es-3* and *Erp-1* loci.

CHAPTER 9

THE GENETIC STRUCTURE OF DIFFERENT MOUSE POPULATIONS

9.1 INTRODUCTION

The genetical structure of a population is herein considered to be the distribution of genes and genotypes over the geographical range of the population (note, this differs from the definition of genetic structure given by Mather, 1973). Expressed more abstractly, the genetic structure is the way in which genetic variation is apportioned within and between subpopulations of the total population. The two extreme types of genetic structure are firstly, the essentially panmictic population without any recognisable subdivisions and secondly, the population divided into discrete subpopulations within which there is little or no polymorphism but with different genes present (often fixed) in separate subpopulations and consequently marked genetic differences between subpopulations (as seen for example in old-world populations of the snail *Rumina decollata*; Selander and Kaufman, 1973). In the first case, genetic variation is uniformly distributed over the whole population whilst in the second, there is negligible variation within subpopulations but extensive variation between subpopulations.

The ecological structure of a population can similarly be defined as the distribution of individuals, and the two extreme forms here are a uniform distribution of individuals across the range of the population or the presence of individuals only in clumps or colonies with wide gaps between inhabited localities. Both the genetic and ecological structures may be affected by the mobility of individuals or gametes (e.g. pollen in plants) and by the breeding system used by the population (e.g. dioecious, monoecious, selfing, asexual).

In this section I will present data that relate to the importance of changes in the ecological structure and the mobility of individuals

in modifying the genetic structure of a population.

The F-statistics discussed in Chapters 2 to 5 will be used to analyse the genetic structures of different mouse populations. From the literature review (Chapter 2), two important predictions can be made about the genetical structure of subdivided natural populations. Firstly, the amount of genetic variation between subpopulations is inversely proportional to the mean subpopulation size and immigration rate. This prediction is most clearly seen in the island model, where (Equation (2.21))

$$F = 1/(1 + 4Nm)$$

and in the linear stepping-stone model (Equation (2.37)), where

$$F_{ST} = 1/(1 + 4N\sqrt{m2k}).$$

Also, in the isolation by distance model over an area population, Equation (2.32),

$$f(o) = 1/(1 - 8\pi \sigma_i^2 d/ \log(2k-k^2))$$

shows that differentiation of subpopulations (as measured by f(o)) is inversely proportional to the population desnity (d) and the dispersal abilities of individuals or gametes (σ_i^2).

This first prediction is very simple for in populations with discrete subpopulations, Nm is the number of migrants per generation and it is intuitively obvious that more immigrants will reduce the genetic differences between a subpopulation and the neighboring subpopulations which supply the immigrants.

There are very serious practical difficulties in estimating N and m in natural populations (Lewontin, 1967). In my attempt to test this prediction about the variation between subpopulations within populations, only crude and relative estimates of Nm can be made. I will take four populations in which house mice are living under different ecological

conditions and on the basis of such ecological data as are available, these four populations will be ranked by the supposed relative values of Nm within their subpopulations. Exact predictions of the mean Nm within subpopulations of each population are not feasible, but relative estimates are possible. The ranking of the populations based on Nm estimates will then be compared with genetic data on the genetical structure of the populations.

The second important prediction about the genetical structure of populations is that for neutral genes, the coefficient of kinship (or correlation in gene frequencies) between two subpopulations should decline exponentially with distance between subpopulations (Equations (2.30),(2.35), (2.36) and (2.38)). A decline in the correlation in gene frequencies between subpopulations implies a corresponding increase in genetic differentiation, which may be measured by a genetic distance statistic. As the geographic distance between randomly chosen subpopulations increases, so should the genetic distance between them. Because the decline in the correlation is exponential (i.e. rapid at first and later more gradual, see Kimura and Weiss, 1964), the genetic distance should rise rapidly and flatten out as the geographic distance increases.

The regression of genetic on geographic distances for non-neutral genes is liable to be variable. If there is uniform selection for an optimum gene frequency over a wide geographic range, genetic distances will be independent of geographic distances. But if gene frequencies are changing clinally with an environmental gradient, then genetic distances may steadily increase with geographic distance.

I will use the data collected from the four mouse populations studied in South Australia and the published data of Petras *et al.* (1969) and Selander, Yang and Hunt (1969) to test the prediction that genetic distances should increase with geographic distances between subpopulations.

TABLE 9.1 Summary of mouse population samples

Sample	Proportion of Population trapped	Total Number Caught	Number Bled	Number of sites trapped	Dates of collection
1970 Plague Population ("Plague")					
Turretfield	negligible	163	161	two	8-10/4/70; 6-7/5/70; 26-29/7/70
Price	negligible	299	57	one	3/6/70
South Kilkerran	negligible	283	65	one	4/6/70
Alford	negligible	328	64	one	4/6/70
Minlaton	negligible	118	60	one	1/7/70
Port Vincent	negligible	140	64	one	2/7/70
Adelaide City Population ("City")					
Adelaide Zoo	moderate	71	68	Several	26/1/71-5/2/71
Pet Shop	almost all	20	17	one	27/4/71-1/5/71; 1-3/6/71
Waite Institute	moderate	17	16	one	28/4/71-1/5/71
University	almost all	41	41	two	1-11/6/71; 21-23/12/71
Oaklands	almost all	45	45	two	20-29/10/71
Westbourne Park	moderate	27	27	two	23-28/9/72
State of South Australia Population ("Country")					
Port Lincoln	negligible	31	30	four	10-12/9/70
Mount Gambier	small	81	76	uncertain	September, 1970
O'Halloran Hill	negligible	84	82	two	22-25/4/71
Wardang Island	negligible	154	144	Several	2-6/5/71
Oraparinna	moderate	32	32	Several	14-19/9/71
Turretfield 1972 Population ("TUR")					
TPA6	almost all	16	14	one	2-4/2/72
TPB6	almost all	12	12	one	2-4/2/72
CRK6	almost all	14	13	one	2-4/2/72
OF6	moderate	14	14	one	2-4/2/72
TPA7	almost all	15	14	one	4-7/3/72
TPB7	almost all	15	15	one	4-7/3/72
CRK7	almost all	29	29	one	4-7/3/72
OF7	moderate	52	51	one	4-7/3/72

MAP 9.1 LEGEND

The sample sites are identified as follows:

P1 = S1 = Turretfield Agricultural Research Centre.

P2 = Price

P3 = South Kilkerran

P4 = Alford

P5 = Minlaton

P6 = Port Vincent

C1 = Adelaide Zoological Gardens

C2 = Pet Shop

C3 = Waite Agricultural Research Institute

C4 = Adelaide University Animal Yards

C5 = Oaklands

C6 = Westbourne Park

S2 = Port Lincoln

S3 = Mount Gambier

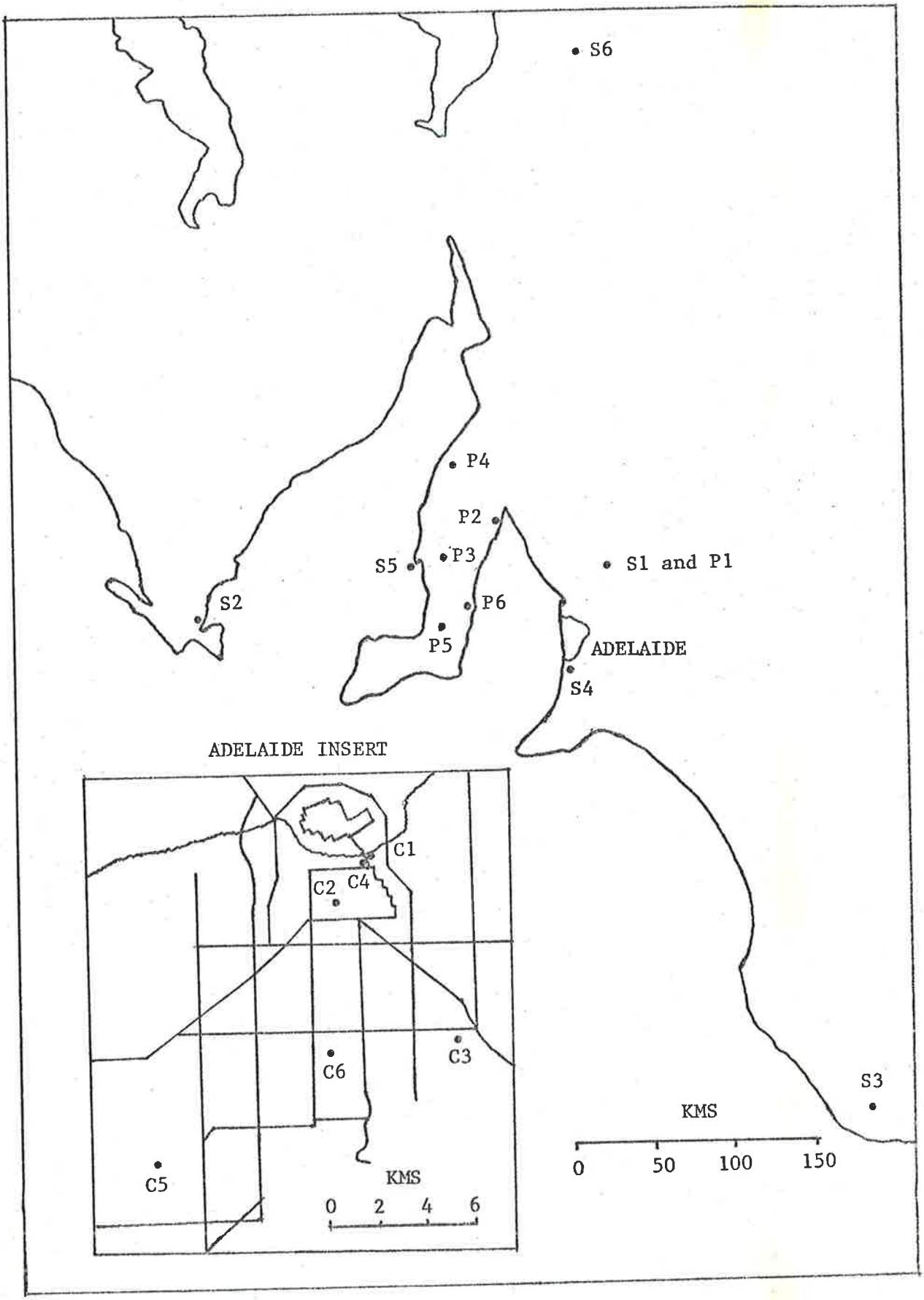
S4 = O'Halloran Hill

S5 = Wardang Island

S6 = Oraparinna

The Turretfield 1972 sites are detailed on Map 10.1.

MAP SHOWING SAMPLE LOCATIONS



MAP 9.1

9.2 DESCRIPTION OF FOUR MOUSE POPULATIONS IN SOUTH AUSTRALIA.

This study was carried out on four mouse populations which are different in geography, habitat, population density and other ecological conditions. The populations (and the 22 subpopulations, each represented by a sample) are given in Table 9.1, which includes some details about each sample. The area covered by each population varies from the few hectares of the Turretfield 1972 population (usually abbreviated to TUR, and described in more detail in Chapter 10) to a large part of the State of South Australia in the "Country" population. The TUR population is represented twice, by two consecutive capture-mark-release censuses, to show how conditions can change in small subpopulations. The locations of the subpopulations sampled are shown on Maps 9.1 and 10.1 (for TUR).

A description of the relevant ecological conditions in each population follows.

9.2.1 The mouse plague of 1970 population

In the autumn and early winter (as late as July in some places) of 1970 there was a serious mouse plague in the grain growing areas north of Adelaide and on Yorke Peninsula. This followed a smaller population peak (with small mouse plagues in some places) in 1969, but without a severe population crash intervening (Newsome, personal communication). Thus the samples from this population were collected at the end of a period when mouse population densities were very much higher than average. During such periods of high densities of mice, many mice can be seen crossing roads at night and trapping at Turretfield yields large numbers of "transient" mice that appear to be entering and leaving the study areas each night (personal observation, and Newsome, personal communication). These observations show that many mice are dispersing across the farmlands at such times, although gene flow need not necessarily be very great because nothing is known about the reproductive performance of these dispersing mice. Despite this lack of precise knowledge, I am confident that relatively large amounts

of gene flow occurred in 1969 and 1970.

Thus Nm was very large for this population.

9.2.2 The City of Adelaide mouse population

Subpopulation samples were relatively difficult to obtain in Adelaide because large colonies of mice are not common in the city as many potentially good habitats (e.g. grain and fodder stores, markets, shops, etc.) are patrolled by pest-control experts who are very efficient at controlling rats and mice. But in the vicinity of animal cages (especially aviaries), poisons are avoided because of the risk to the captive animals. Because trapping is time-consuming and inefficient, such places often maintain mouse colonies which thrive on the food and shelter provided. These subpopulations fluctuate greatly in numbers because when mice become so prevalent that serious damage occurs to food stocks, a mouse destruction campaign is usually carried out to attempt to exterminate them. Such control measures as are taken (trapping, poisoning, flooding mouse burrows) usually effect a drastic reduction in mouse numbers, occasionally even leading to a temporary extinction, and the colony may take many months (or even years) to reach high numbers again. Hence the effective size of these subpopulations is rather small.

Movement between city subpopulations will be restricted by areas of unsuitable habitat (roads, lawns, mouse-proof buildings) and a high density of predators, especially cats and dogs. Colonisation of new areas of suitable habitat and recolonisation of sites where a previous colony was exterminated, will be by occasional dispersing mice that manage to evade predators and cross the inhospitable surroundings and by mice that "hitch-hike" their way to the site in hay bales and other food supplies.

The presence of a low number of vagrant mice in the city is confirmed by three observations. Firstly, city householders sometimes find a mouse or two raiding foodstuffs in the kitchen and causing a nuisance until

caught or poisoned. Secondly, domestic cats occasionally bring home dead mice that they have caught, and thirdly, in three years I have caught five wild mice in the lowest three floors of the R.A. Fisher Building which most certainly were not escapees from the mouse stocks kept on the sixth floor. These mice are probably emigrants from successful mouse colonies with a reproductive surplus or are dispersing from areas that have ceased to be habitable and are searching for fresh food and shelter.

Newsome (1971) has documented the occurrence of mice in South Australian haystacks. Hay-bales brought into the city for fodder and bedding for animals are likely to contain mice from the country. Old sacks and partly empty bags and cartons of food sometimes contain mice and if these are moved around the city, then the mice will be carried along. All these observations show that mice can move into and around the city, but probably at a low rate.

Hence for the city, Nm will probably be small.

9.2.3 Country population in South Australia

Five mouse subpopulations in widely separated country locations were sampled to assess the amount of genetic variation between different parts of the State. The sample from Turretfield collected in the 1970 plague was added to these five to represent the mice from the wheatfields north of Adelaide.

Movement of mice over long distances will usually occur inadvertantly in vehicles carrying food and fodder and may not be of common occurrence. Wardang Island is the most isolated subpopulation that was sampled for it is accessible only by boat (and recently, by plane) and although inhabited by Europeans since the last century (a small mine was there, now a small tourist resort), only relatively few mice are likely to have made their way across in supplies for the inhabitants or to have escaped from wrecks of grain ships on the island in the last century (Mr. G. Price,

personal communication). For the other subpopulations sampled, isolation by distance would reduce the effective amount of gene flow between them. Thus m is probably very small indeed for these country subpopulations and Nm will be small regardless of the size of N.

9.2.4 The Turretfield intensive study, 1972

This study will be dealt with in detail in Chapter 10. Here I will deal with the data for censuses 6 and 7 of three mouse colonies established around feedstations (denoted TPA, TPB and CRK) and a fourth subpopulation in a fenced-off and overgrown study area known as the old field (denoted OF) (see Map 10.1). These four subpopulations are small (although CRK and OF expanded between censuses 6 and 7), but genetic evidence discussed in Chapter 10 suggests that there is a high rate of immigration into each of these subpopulations. Hence Nm is probably smaller than in the 1970 plague, but larger than in the city or country populations.

9.2.5 Description of the samples

A few details of each subpopulation sample are presented in Table 9.1. For each sample, an attempt was made to indicate roughly the proportion of the local subpopulation that was collected because these proportions are heterogeneous and the statistical methods used to analyse the data should differ according to the proportion of each subpopulation included in the sample. For all of the plague samples, only a very small fraction of the thousands of mice at each sample location were taken, but, by comparison, three of the city samples were obtained by trapping until no more mice could be caught and in the three feed-station subpopulations at Turretfield most of the mice were caught on a trap-release study.

For each sample, the number of distinct sites at which mice were obtained is indicated in Table 9.1. Each site was a clearly defined area

TABLE 9.2 Genotype numbers at the *Es-3* locus.

Sample	*cc*	*bc*	*bb*	*dd*	*dc*	*db*	Total	Frequency of *Es-3*^{c}
Turretfield	64	80	17				161	$.646^{+}_{-}.025$
Price	28	25	4				57	$.711^{+}_{-}.041$
South Kilkerran	36	25	4				65	$.746^{+}_{-}.038$
Alford	39	22	3				64	$.781^{+}_{-}.036$
Minlaton	32	23	5				60	$.725^{+}_{-}.042$
Port Vincent	36	24	4				64	$.750^{+}_{-}.038$
Plague Total	235	199	37				471	$.710^{+}_{-}.015$
Adelaide Zoo	6	24	38				68	$.265^{+}_{-}.040$
Pet Shop	3	1	0	6	7	0	17	$.412^{+}_{-}.086$
Waite Institute	3	11	2				16	$.531^{+}_{-}.069$
University	16	17	8				41	$.598^{+}_{-}.058$
Oaklands	18	17	10				45	$.589^{+}_{-}.057$
Westbourne Park	0	7	19				26	$.135^{+}_{-}.043$
City Total	46	77	77	6	7	0	213	$.413^{+}_{-}.024$
Port Lincoln	1	0	29				30	$.033^{+}_{-}.033$
Mt. Gambier	14	25	18				57	$.465^{+}_{-}.049$
O'Halloran Hill	28	40	12	0	2	0	82	$.598^{+}_{-}.037$
Wardang Island	105	29	10				144	$.830^{+}_{-}.025$
Oraparinna	8	16	5	0	1	2	32	$.516^{+}_{-}.060$
Country Total*	220	190	91	0	3	2	506	$.625^{+}_{-}.015$
TPA6	6	7	0				13	$.731^{+}_{-}.069$
TPB6	5	6	1				12	$.667^{+}_{-}.090$
CRK6	5	7	1				13	$.654^{+}_{-}.084$
OF6	7	6	1				14	$.714^{+}_{-}.083$
TOT6	23	26	3				52	$.692^{+}_{-}.045$
TPA7	6	8	0				14	$.714^{+}_{-}.066$
TPB7	4	8	3				15	$.533^{+}_{-}.088$
CRK7	13	14	2				29	$.690^{+}_{-}.057$
OF7	29	19	3				51	$.755^{+}_{-}.042$
TOT7	52	49	8				109	$.702^{+}_{-}.031$

* Includes Turretfield.

which appeared distinct from other sites either by distance or because mice did not appear to go from one site to another even though the distance was only 10 to 20 metres. In some cases the exact number of sites is unknown, but is more than two or three: these cases are indicated as "several sites". When possible, the genetic data from different sites were kept separate so that the gene frequencies could be tested for homogeneity across sites. The only case where heterogeneity was evident was at the Adelaide Zoo, where mice under the bird cages (north to eastern part of the Zoo) differed significantly from mice caught in several large animal cages (west to south part of the Zoo) at the *Es-3* locus ($\chi^2_1 = 16.5$). It was not possible to test for homogeneity at any of the other loci scored and all data were pooled to give a sample that is intended to represent the Zoo mice as a whole.

The mice caught in the store-room of the City Pet Shop appeared to be fully wild and none had the coat-colour variants (white, black, brown, etc.) typical of the tame mice sold in the Pet Shop. When these mice were caught, I was confident that they were wild mice and not escapees from the tame stocks kept for sale.

9.3 RESULTS

9.3.1 Genotype and Phenotype numbers

The genotype numbers for the *Es-3*, *Hbb*, *Erp-1* and *Es-1* loci and phenotype numbers for the *A* locus are given in Tables 9.2 to 9.6 respectively.

At the *Es-3* locus (Table 9.2), the $Es\text{-}3^c$ allele is usually most frequent except in the city. $Es\text{-}3^d$ is a rare allele that occurs sporadically and was common only in the small sample from the store-room of the Pet Shop. In this case, the high frequency of $Es\text{-}3^d$ might be explained by founder effect, or random sampling drift in a small subpopulation, or

TABLE 9.3 Genotype numbers of the *Hbb* locus.

Sample	*ss*	*sd*	*dd*	*d?*	Total	Frequency of Hbb^S
Turretfield	97(82)*	41	3	8	(126)*	$.813^{*+}_{-}.024$
Price	41(39)*	14	1	1	(54)*	$.852^{*+}_{-}.034$
South Kilkerran	45	19	1	0	65	$.838^{+}_{-}.031$
Alford	45	17	2	0	64	$.836^{+}_{-}.033$
Minlaton	46	13	1	0	60	$.875^{+}_{-}.030$
Port Vincent	44	20	0	0	64	$.844^{+}_{-}.029$
Plague Total	301*	124	8		433*	$.838^{+}_{-}.013$
Adelaide Zoo	66	0	0	0	66	1.000
Pet Shop	16	1	0	0	17	$.971^{+}_{-}.029$
Waite Institute	7	8	1	0	16	$.688^{+}_{-}.075$
University	37	4	0	0	41	$.951^{+}_{-}.023$
Oaklands	21	17	7	0	45	$.656^{+}_{-}.054$
Westbourne Park	21	4	2	0	27	$.852^{+}_{-}.057$
City Total	168	34	10		212	$.873^{+}_{-}.016$
Port Lincoln	30	0	0	0	30	1.000
Mt. Gambier	52	21	3	0	76	$.822^{+}_{-}.032$
O'Halloran Hill	36	35	11	0	82	$.652^{+}_{-}.038$
Wardang Island	144	0	0	0	144	1.000
Oraparinna	26(22)*	3	2	1	(27)*	$.870^{*+}_{-}.056$
Country Total**	366*	100	19		485*	$.858^{+}_{-}.011$
TPA6	5	8	0	0	13	$.692^{+}_{-}.067$
TPB6	10	2	0	0	12	$.917^{+}_{-}.054$
CRK6	5	7	1	0	13	$.654^{+}_{-}.084$
OF6	7	7	0	0	14	$.750^{+}_{-}.067$
TOT6	27	24	1		52	$.750^{+}_{-}.042$
TPA7	10	4	0	0	14	$.857^{+}_{-}.060$
TPB7	12	2	0	0	14	$.929^{+}_{-}.047$
CRK7	15	14	0	0	29	$.759^{+}_{-}.046$
OF7	25	23	3	0	51	$.716^{+}_{-}.042$
TOT7	62	43	3		108	$.773^{+}_{-}.029$

* Corrected for number of *d?* animals.

** Includes Turretfield.

TABLE 9.4 Genotype numbers at the *Erp-1* locus.

Sample	*aa*	*ab*	*bb*	Total	Frequency of *Erp-1*[a]
Turretfield	113	2	0	115	$.991^{+}_{-}.006$
Price	37	0	0	37	1.000
South Kilkerran	41	4	0	45	$.956^{+}_{-}.021$
Alford	46	1	0	47	$.989^{+}_{-}.011$
Minlaton	50	1	0	51	$.990^{+}_{-}.010$
Port Vincent	33	2	0	35	$.971^{+}_{-}.020$
Plague Total	320	10	0	330	$.985^{+}_{-}.005$
Adelaide Zoo	36	0	0	36	1.000
Pet Shop	16	0	0	16	1.000
Waite Institute	10	2	0	12	$.917^{+}_{-}.054$
University	40	0	0	40	1.000
Oaklands	44	0	0	44	1.000
Westbourne Park	27	0	0	27	1.000
City Total	173	2	0	175	$.994^{+}_{-}.004$
Port Lincoln	30	0	0	30	1.000
Mt. Gambier	52	0	0	52	1.000
O'Halloran Hill	58	3	0	61	$.975^{+}_{-}.014$
Wardang Island	61	3	0	64	$.977^{+}_{-}.013$
Oraparinna	13	0	0	13	1.000
Country Total*	327	8	0	335	$.988^{+}_{-}.006$
TPA6	14	0	0	14	1.000
TPB6	12	0	0	12	1.000
CRK6	13	0	0	13	1.000
OF6	14	0	0	14	1.000
TOT6	53	0	0	53	1.000
TPA7	13	1	0	14	$.964^{+}_{-}.034$
TPB7	14	0	0	14	1.000
CRK7	29	0	0	29	1.000
OF7	46	5	0	51	$.951^{+}_{-}.021$
TOT7	102	6	0	108	$.972^{+}_{-}.011$

* Includes Turretfield

TABLE 9.5 Genotype numbers at the *Es-1* locus.

Sample	*bb*	*ab*	*aa*	Total	Frequency of *Es-1*[b]
Turretfield	127	10	0	137	$.964^{+}_{-}.011$
Price	55	2	0	57	$.982^{+}_{-}.012$
South Kilkerran	61	3	1	65	$.962^{+}_{-}.020$
Alford	57	6	1	64	$.938^{+}_{-}.023$
Minlaton	57	2	0	59	$.983^{+}_{-}.012$
Port Vincent	64	0	0	64	1.000
Plague Total	421	23	2	446	$.970^{+}_{-}.006$
Adelaide Zoo	63	0	0	63	1.000
Pet Shop	17	0	0	17	1.000
Waite Institute	16	0	0	16	1.000
University	39	0	0	39	1.000
Oaklands	45	0	0	45	1.000
Westbourne Park	26	1	0	27	$.981^{+}_{-}.018$
City Total	206	1	0	207	$.998^{+}_{-}.002$
Port Lincoln	28	1	0	29	$.983^{+}_{-}.017$
Mt. Gambier	58	4	0	62	$.968^{+}_{-}.016$
O'Halloran Hill	82	1	0	83	$.994^{+}_{-}.006$
Wardang Island	118	1	0	119	$.996^{+}_{-}.004$
Oraparinna	28	4	0	32	$.938^{+}_{-}.029$
Country total*	441	21	0	462	$.977^{+}_{-}.005$
TPA6	8	6	0	14	$.786^{+}_{-}.066$
TPB6	9	3	0	12	$.875^{+}_{-}.063$
CRK6	10	2	0	12	$.917^{+}_{-}.054$
OF6	12	2	0	14	$.929^{+}_{-}.047$
TOT6	39	13	0	52	$.875^{+}_{-}.032$
TPA7	11	3	0	14	$.893^{+}_{-}.055$
TPB7	12	3	0	15	$.900^{+}_{-}.052$
CRK7	20	9	0	29	$.845^{+}_{-}.043$
OF7	46	5	0	51	$.951^{+}_{-}.021$
TOT7	89	20	0	109	$.908^{+}_{-}.020$

* Includes Turretfield

TABLE 9.6 Phenotype numbers at the *Agouti* locus.

Sample	A^w.	A^+A^+	Total	frequency A^+ allele			
				F=0.0	F=.0986	F=-.071	F=-.199
Turretfield	149	14	163	.2931	.2588		
Price	288	11	299	.1918		.2216	
South Kilkerran	257	26	283	.3031		.3280	
Alford	315	13	328	.1991		.2285	
Minlaton	111	7	118	.2436		.2710	
Port Vincent	128	12	140	.2928		.3181	
Plague Total	1248	83	1331	.245			
Adelaide Zoo	4	67	71	.9714	.9699		
Pet Shop	2	18	20	.9487		.9505	
Waite Institute	14	3	17	.4201		.4405	
University	27	14	41	.5843	.5632		
Oaklands	40	5	45	.3333	.3006		
Westbourne Park	20	7	27	.5092	.4844		
City Total	107	114	221	.669			
Port Lincoln	8	23	31	.8614	.8542		
Mt. Gambier	0	81	81	1.0000	1.0000		
O'Halloran Hill	68	16	84	.4364	.4082		
Wardang Island	141	13	154	.2905	.2562		
Oraparinna	20	12	32	.6124	.5926		
Country Total*	386	159	545	.462			
TPA6	16	0	16	.0000			.1662
TPB6	6	6	12	.7071			.7341
CRK6	14	0	14	.0000			.1662
OF6	13	1	14	.2673			.3409
TOT6	49	7	56	.218			
TPA7	15	0	15	.0000			.1662
TPB7	10	5	15	.5774			.6168
CRK7	27	2	29	.2626			.3369
OF7	51	1	52	.1387			.2346
TOT7	103	8	111	.212			

* Includes Turretfield.

TABLE 9.7 Sample fixation Indices (F_{IS})

Sample	*Hbb*	*Es-3*	*Es-1*	*Erp-1*
Turretfield	-.072* ± .079	-.086 ± .077	-.038 ± .012	-.009 ± .006
Price	-.027* ± .128	-.066 ± .128	-.018 ± .013	
South Kilkerran	-.079 ± .103	-.015 ± .123	.376 ± .284	-.047 ± .023
Alford	.032 ± .131	-.006 ± .124	.200 ± .211	-.011 ± .011
Minlaton	.010 ± .132	.039 ± .131	-.017 ± .012	-.010 ± .010
Port Vincent	-.185 ± .041	.000 ± .125	-	-.029 ± .021
Adelaide Zoo		.093 ± .126	-	-
Pet Shop	-.030 ± .030	.029† ± .243	-	-
Waite Institute	-.164 ± .229	-.380 ± .230	-	-.091 ± .064
University	-.051 ± .026	.138 ± .156	-	-
Oaklands	.164 ± .151	.220 ± .147	-	-
Westbourne Park	.413 ± .242	-.156 ± .058	-.019 ± .019	-
Port Lincoln		1.00 ± .000	-.018 ± .018	-
Mt. Gambier	.054 ± .122	.119 ± .132	-.033 ± .017	-
O'Halloran Hill	.059 ± .111	-.065† ± .110	-.006 ± .006	-.025 ± .015
Wardang Island		.287 ± .101	-.004 ± .004	-.024 ± .014
Oraparinna	.508* ± .247	-.064† ± .176	-.067 ± .033	-
TPA6	-.444 ± .141	-.368 ± .129	-.273 ± .107	-
TPB6	-.091 ± .064	-.125 ± .275	-.143 ± .082	-
CRK6	-.190 ± .257	-.190 ± .257	-.091 ± .064	-
OF6	-.333 ± .119	-.050 ± .260	-.077 ± .054	-
TPA7	-.167 ± .082	-.400 ± .130	-.120 ± .069	-.037 ± .037
TPB7	-.077 ± .054	-.071 ± .257	-.111 ± .064	-
CRK7	-.318 ± .081	-.128 ± .174	-.184 ± .060	-
OF7	-.108 ± .131	-.007 ± .139	-.052 ± .023	-.052 ± .023

* Corrected for number of $Hbb^{d?}$ animals.

† With $Es\text{-}3^{d}$ pooled with $Es\text{-}3^{b}$.

perhaps there was an escapee from the tame mice in the shop. As nothing is known about the frequency of *Es-3*^{d} in tame mice, there is no evidence for or against the third possibility.

The *Hbb*^{s}*Hbb*^{d} heterozygotes could not be scored reliably until Dr. Petras' suggestion (Chapter7), but by this time some haemolysates already scored as *Hbb*^{d}. had become exhausted and it was impossible to rescore them. In three samples, the number of *Hbb*^{s}*Hbb*^{s} individuals was reduced to compensate for the proportion of *Hbb*^{d}. animals that could not be rescored for their exact genotype. This meant discarding some data, but this procedure should eliminate the bias in genotype frequencies that would otherwise occur.

Hbb^{s} is the common allele and appears to be fixed in three of the subpopulations sampled (Table 9.3).

In some samples, fewer mice are scored for the *Erp-1* and *Es-1* loci because the techniques for scoring these loci were not developed until late in the project and some haemolysate and serum samples had become exhausted. The *Erp-1* locus (Table 9.4) is monomorphic in several samples and *Erp-1*^{b} is a rare gene, occurring with a frequency about 1% over all samples. *Es-1*^{a} is also a rare gene, particularly in the city samples, with a frequency less than 3% in most samples except the Turretfield 1972 samples (Table 9.5).

At the *A* locus there is a wide range of phenotype frequencies, from all A^{w}. in TPA to all $A^{+}A^{+}$ at Mount Gambier. The estimated frequency of the recessive A^{+} allele is given in Table 9.6, twice for each sample corresponding to two of four values for the fixation index (F_S) applicable to the sample. The estimation of these fixation indices is given in the next section.

9.3.2 Sample Fixation Indices

For each locus with codominant alleles Table 9.7 gives the estimated

fixation index (F_{IS}) and standard error (from Rasmussen, 1964; and Brown, 1970). Where one allele is fixed in a sample, no estimate of F_{IS} is possible and this event is indicated by a dash. At the *Es-3* locus, the rare *Es-3*^{d} allele was treated as though it were *Es-3*^{b}, to simplify the analysis. Because the data on *Erp-1* and *Es-1* are incomplete and the rarity of the less common allele at each locus might bias estimates of F_{IS} because only one homozygote occurs, I decided to deal only with the data on *Hbb* and *Es-3* in a detailed analysis.

F_{IS} varies considerably between different samples and two questions can be asked about this variation:

(1) is there any difference in F_{IS} values between *Es-3* and *Hbb?*

(2) can we classify the samples in any meaingful way which will throw light on the cause of some, at least, of this variation?

In a preliminary attempt to answer question (1), the correlation between F_{IS}(*Es-3*) and F_{IS}(*Hbb*) was calculated over the 22 samples in which valid estimates of F_{IS} were obtained for both loci. The correlation was .37, which is not significantly different from no correlation at all.

The theory relating F_{IS} to the number of distinct subpopulations included in a conglomerate subpopulation (Chapter 3.11) is useful in answering both questions. The theory predicts that a single subpopulation should have a negative F_{IS} and that with an increasing number of distinct subpopulations included in a conglomerate subpopulation, F_{IS} will become more positive.

The samples collected from one site may represent single subpopulations and the samples from several sites may represent conglomerate subpopulations. Hence the samples can be split into two classes: those from one site and those from two or more sites. The samples from one site can be further subdivided into those where the existence of one subpopulation is known from other data (Turretfield, 1972 populations, Chapter 10)and those where

TABLE 9.8 SUMMARY OF ANALYSIS OF DIFFERENCES BETWEEN LOCI AND BETWEEN FOUR CLASSES OF SAMPLES FOR F_{IS}

Loci	CLASSES			
	Plague + City	One Site TUR6	TUR7	Two or more sites
	n = 7	n = 4	n = 4	n = 9
Es-3	-0.067 ± 0.060	-0.181 ± 0.066	-0.150 ± 0.085	0.064 ± 0.069
Hbb	-0.078 ± 0.034	-0.294 *± 0.073	-0.184 *± 0.056	0.165 ± 0.073
t-test for difference $F_{IS}(Es\text{-}3)-F_{IS}(Hbb)$	+0.16	+1.16	+0.33	-1.02
F_{IS} after pooling both loci	-0.071 ± 0.035	-0.232 **± 0.050	-0.164 *± 0.049	0.099 ± 0.049
t-test of pairwise comparisons	2.68 *	1.56	0.97	2.72 *	4.18 **	3.30 **
Pooling TUR6 + TUR7		-0.199 **± 0.035		

* Probability < 5%

** Probability < 1%

no data are available on the actual number of mouse subpopulations sampled. Because censuses 6 and 7 at Turretfield are consecutive censuses, there may be correlations between F_{IS} values from the same site and hence the analysis of these two censuses will be kept separate until the end.

The calculations of the mean F_{IS} (denoted $\hat{F}_{IS}$) for a set of samples and its standard error is complicated because each sample F_{IS_i} is weighted by $q_i(1-q_i)$ where F_{IS_i} and q_i are the sample fixation index and gene frequency, respectively. Denoting the weighting for the ith sample by w_i, then $\hat{F}_{IS} = \sum_{i=1}^{n} w_i F_{IS_i} / \sum_{i=1}^{n} w_i$ for n subpopulation samples. The variance (s^2) of F_{IS_i} values is

$$s^2 = \left(\sum_{i=1}^{n} w_i F_{IS_i}^2 - \left(\sum_{i=1}^{n} w_i\right)\hat{F}_{IS}^2\right) \Big/ \left(\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 \Big/ \sum_{i=1}^{n} w_i\right) \qquad \ldots\ldots(9.1)$$

$$= A/B, \text{ say.}$$

The standard error of $\hat{F}_{IS}$ is

$$SE = \sqrt{s^2/n}$$

The number of degrees of freedom for s^2 can be found from

$$dof = n\left(1 - \sum_{i=1}^{n} w_i^2 \Big/ \left(\sum_{i=1}^{n} w_i\right)^2\right)$$

which is usually only slightly less than (n-1). Two $\hat{F}_{IS}$ values (denoted $\hat{F}_{IS1}$ and $\hat{F}_{IS2}$) with variances $s_1^2 = A_1/B_1$ and $s_2^2 = A_2/B_2$ that are not heterogeneous may be compared by a t-test by calculating

$$s_{12}^2 = (A_1 + A_2)/(B_1 + B_2)$$

and $t_{(dof_1 + dof_2)} = (\hat{F}_{IS1} - \hat{F}_{IS2})/(s_{12}(1/n_1 + 1/n_2))$

where dof_i and n_i are the degrees of freedom and sample sizes of the ith $\hat{F}_{IS}$.

Table 9.8 lists the $\hat{F}_{IS} \pm$ SE estimates for four classes of samples: three classes of samples from one site (plague + city, TUR6 and TUR7) and one class for the samples from two or more sites. For all classes, the

differences between $\hat{F}_{IS}$(*Es-3*) and $\hat{F}_{IS}$(*Hbb*) are not significant by t-tests and so the data were pooled over loci within each class to obtain an overall $\hat{F}_{IS} \pm$ SE. Pairwise comparisons between the four classes reveal that Turretfield 6 and 7 are not different, the "plague + city" class is different from Turretfield 6 but not 7 and that F_{IS} from two or more sites is significantly different from any of the $\hat{F}_{IS}$ values from one site.

These results agree with the theory of F_{IS} discussed above because $\hat{F}_{IS} > 0$ in the samples from two or more sites and $\hat{F}_{IS} << 0$ in the samples known to be from single subpopulations. The significant variation in $\hat{F}_{IS}$ values between classes shows that $\hat{F}_{IS} \neq 0$ in general, so that when gene frequencies are estimated at the *agouti* locus each sample should be given the F_{IS} value appropriate to the class of sample to which it belongs. Since the Turretfield samples did not differ between the censuses, the data were pooled over censuses 6 and 7 to yield $\hat{F}_{IS} = -.199 \pm .035$ as the estimated fixation index in those samples.

This analysis demonstrates an important fact: namely, that the estimate of F_{IS} in a sample can be strongly affected by the way in which the sample is collected.

9.3.3 Genetic structure: analysis by F-statistics

9.3.3.1 Methods

The three F-statistics (Wright, 1965) F_{IS}, F_{ST} and F_{IT} can be used to analyse the genetic structure of a population.

There are two sources of heterogeneity that make it impossible to have a completely rigorous estimation procedure for the F-statistics in all populations. Firstly, there is the heterogeneity of $\hat{F}_{IS}$ values due to variation in the number of sites contributing to each sample. Secondly there is variation in the efficiency with which each subpopulation was sampled; some samples fulfil the conditions of binomial sampling with replacement, others represent almost a complete enumeration of a sub-

population without sampling error; and other samples are intermediate.

There is little that can be done about the first problem except to note that since F_{IS} should be negative within single subpopulations of house-mice, if the estimate of F_{IS} for a mouse population is positive, then most samples have presumably come from conglomerate subpopulations. F_{IT} should not be greatly affected by sampling from conglomerate subpopulations; consequently, because of the fixed relations between F_{IT}, F_{ST} and F_{IS}, F_{ST} between single subpopulations will be underestimated if F_{IS} is positive in samples.

To partially overcome the second problem, all the data were analysed in two ways by making different assumptions about the sampling procedures. Firstly, "sample" F-statistics were calculated assuming that each subpopulation in the population was completely censused and there was no sampling error in estimating F_{IS} and F_{ST}. This could lead to an underestimate of F_{IS} because binomial sampling reduces F_{IS} (Equation (4.20)); and probably an underestimate of F_{ST} because no Gaussian correction (n/(n-1)) was made in estimating the gene frequency variance between samples. Secondly, "population" F-statistics were calculated assuming that only a small proportion of each of a small proportion of all possible subpopulations was sampled (i.e. binomial sampling). Thus each sample F_{IS} was converted to a population estimate using Equation (4.24) on the assumption of binomial sampling. The estimate of F_{ST} not only had a Gaussian correction applied but also had the mean sampling variance subtracted (Nei and Imaizumi,1966). Also, whereas in the "sample" F-statistics, each sample was given equal weighting, in the "population" F-statistics each sample was weighted by sample size.

The methods of calculation are as follows, assuming the ith of n samples to be of size N_i with gene frequency q_i and heterozygote frequency, H_i.

(a) "Sample" F-statistics:

$$F_{IS_i} = 1 - H_i/2q_i(1-q_i)$$

$$F_{IS} = \sum_{i=1}^{n} q_i(1-q_i)\, F_{IS_i} / \sum_{i=1}^{n} q_i(1-q_i)$$

$$\bar{q} = \frac{1}{n} \sum_{i=1}^{n} q_i$$

$$s_q^2 = \frac{1}{n} \sum_{i=1}^{n} (q_i-\bar{q})^2$$

$$F_{ST} = s_q^2 /\bar{q}(1-\bar{q})$$

$$\bar{H} = \frac{1}{n} \sum_{i=1}^{n} H_i$$

$$F_{IT} = 1 - \bar{H}/2\bar{q}(1-\bar{q})$$

These three F-statistics obey the relation $(1-F_{IT}) = (1-F_{IS})(1-F_{ST})$.

(b) Population F-statistics:

$$F_i = 1 - H_i/2q_i(1-q_i)$$

$$F_{IS_i} = ((2N_i-1)F_i + 1)/(2N_i + F_i - 1) \quad \text{(from Equation (4.24))}$$

$$F_{IS} = \sum_{i=1}^{n} N_i q_i(1-q_i) F_{IS_i} / \sum_{i=1}^{n} N_i q_i(1-q_i) \qquad \ldots\ldots(9.2)$$

To calculate the variance between samples, taking account of the Gaussian correction and subtracting the mean sampling variance, it is convenient to use the method of calculating the between samples component of variance from an analysis of variance. The mean effective sample number of genes ($2N_o$) is

$$2N_o = \left(\sum_{i=1}^{n} 2N_i - \sum_{i=1}^{n} (2N_i)^2 / \sum_{i=1}^{n} 2N_i \right)/(n-1) \qquad \ldots\ldots(9.3)$$

and

$$\bar{q} = \sum_{i=1}^{n} N_i q_i / \sum_{i=1}^{n} N_i \qquad \ldots\ldots(9.4)$$

The weighted observed variance between samples is

$$s_o^2 = \sum_{i=1}^{n} 2N_i (q_i-\bar{q})^2 / 2N_o n \qquad \ldots\ldots(9.5)$$

and the weighted mean sampling error is

$$\sigma_s^2 = \sum_{i=1}^{n} 2N_i q_i(1-q_i)/2N_o\left(\sum_{i=1}^{n}(2N_i-1)\right) \qquad \ldots\ldots(9.6)$$

The estimate of the between samples component of variation is

$$s_q^2 = n\, s_o^2/(n-1) - \sigma_s^2 \qquad \ldots\ldots(9.7)$$

and as before,

$$F_{ST} = s_q^2/\bar{q}(1-\bar{q}) \quad .$$

Wright (1943b) points out that this method of estimating the variance between populations can be biassed, especially when s_o^2 is large and one allele is rare, and he provides a correction. His equations deal with the simpler case of all N_i=N, and hence

$$s_o^2 = \sum_{i=1}^{n}(q_i-\bar{q})^2/n$$

and

$$\sigma_s^2 = \sum_{i=1}^{n} q_i(1-q_i)/n\,(2N-1)$$

Wright's equation for estimating s_q^2 is

$$s_q^2 = \left(n\, s_o/(n-1+s_o/\bar{q}(1-\bar{q}))\right) - \sigma_s^2 \qquad \ldots\ldots(9.8)$$

Assuming this equation was valid when s_o^2 and σ_s^2 were calculated from samples weighted by N_i, s_q^2 was also calculated by this equation to test whether or not Wright's correction makes a significant difference to the estimates of F_{ST} obtained in this way.

"Population" F_{IT} was calculated from the relation $F_{IT} = F_{ST}+F_{IS}(1-F_{ST})$ where F_{ST} and F_{IS} were obtained as above.

A third estimate of F_{IT} for each population was obtained by pooling all the genotype data for that population and calculating the fixation index: this was called the "pooled sample" estimate of F_{IT}. (F_{IS} standard errors from Equation (9.1) and F_{IT} (pooled sample) standard errors from Equation (4.25).).

TABLE 9.9 F-Statistics for 5 mouse populations: *Es-3* locus

		Plague	TUR6	TUR7	City	Country
F_{IS}(sample)		$\frac{-.025}{+.019}$	$\frac{-.181}{+.066}$	$\frac{-.150}{+.085}$	$\frac{-.001}{+.095}$	$\frac{+.044}{-.094}$
F_{IS}(population estimate)		$\frac{-.033}{+.022}$	$\frac{-.141}{+.067}$	$\frac{-.083}{+.071}$	$\frac{+.084}{-.085}$	$\frac{+.040}{-.082}$
F_{ST}(sample)		.009	.005	.032	.123	.239
F_{ST}(population estimate)	(a)	.009	-.026	.019	.146	.197
	(b)	.009	-.026	.019	.142	.190
F_{IT}(pooled sample)		$\frac{-.026}{+.046}$	$\frac{-.174}{+.125}$	$\frac{-.074}{+.092}$	$\frac{+.187}{-.068}$	$\frac{+.186}{-.045}$
F_{IT}(sample)		-.016	-.175	-.113	.121	.273
F_{IT}(population estimate)		-.024	-.170	-.062	.218	.229

(a) s_q^2 estimated from ordinary between samples component of variation.

(b) s_q^2 with Wright's correction applied.

TABLE 9.10 F-Statistics for 5 mouse populations: *Hbb* locus.

		Plague	TUR6	TUR7	City	Country
F_{IS}(sample)		$\frac{-.055}{+.032}$	$\frac{-.294}{+.073}$	$\frac{-.184}{+.056}$	$\frac{+.079}{-.103}$	$\frac{+.106}{-.092}$
F_{IS}(population estimate)		$\frac{-.053}{+.030}$	$\frac{-.261}{+.073}$	$\frac{-.157}{+.062}$	$\frac{+.139}{-.090}$	$\frac{+.042}{-.067}$
F_{ST}(sample)		.003	.054	.046	.148	.119
F_{ST}(population estimate)	(a)	-.003	.041	.031	.209	.152
	(b)	-.003	.040	.030	.201	.148
F_{IT}(pooled sample)		$\frac{-.057}{+.043}$	$\frac{-.231}{+.100}$	$\frac{-.135}{+.081}$	$\frac{+.279}{-.092}$	$\frac{+.155}{-.055}$
F_{IT}(sample)		-.053	-.224	-.131	.214	.214
F_{IT}(Population estimate)		-.056	-.210	-.121	.319	.189

(a) s_q^2 estimated from ordinary between samples component of variation.

(b) s_q^2 with Wright's correction applied.

TABLE 9.11 F-Statistics for 5 mouse populations: *Hbb+Es-3* pooled.

	Plague	TUR6	TUR7	City	Country
F_{IS}(sample)	$\frac{-.037}{+.018}$	$\frac{-.232}{+.050}$	$\frac{-.164}{+.049}$	$\frac{+.026}{-.066}$	$\frac{+.067}{-.063}$
F_{IS}(population estimate)	$\frac{-.040}{+.017}$	$\frac{-.196}{+.051}$	$\frac{-.116}{+.045}$	$\frac{+.101}{-.053}$	$\frac{+.041}{-.051}$
F_{ST}(sample)	$\frac{+.006}{-.003}$	$\frac{+.030}{-.025}$	$\frac{+.039}{-.007}$	$\frac{+.136}{-.013}$	$\frac{+.179}{-.060}$
F_{ST}(population estimate)	$\frac{+.003}{-.006}$	$\frac{+.008}{-.034}$	$\frac{+.025}{-.006}$	$\frac{+.178}{-.032}$	$\frac{+.175}{-.023}$
F_{IT}(pooled sample)	$\frac{-.041}{+.031}$	$\frac{-.202}{+.080}$	$\frac{-.105}{+.061}$	$\frac{+.232}{-.057}$	$\frac{+.171}{-.035}$
F_{IT}(sample)	$\frac{-.035}{+.019}$	$\frac{-.200}{+.025}$	$\frac{-.122}{+.009}$	$\frac{+.168}{-.047}$	$\frac{+.244}{-.030}$
F_{IT}(population estimate)	$\frac{-.040}{+.016}$	$\frac{-.190}{+.020}$	$\frac{-.092}{+.030}$	$\frac{+.269}{-.051}$	$\frac{+.209}{-.020}$

TABLE 9.12 Analysis of Variance of $\sqrt{F_{ST}}$, with the *Es-3* and *Hbb* loci treated as replicates within each population.

Source of Variation	Sums of Squares	Degrees of freedom	Mean Square	$F_{4,5}$
Between	.16956	4	.04239	
Within	.02550	5	.00510	8.31 P < 5%
Total	.19506	9		

TABLE 9.13 Analysis of variance of "population" estimates of F_{IT} with the *Es-3* and *Hbb* loci treated as replicates within each population. Z-transform used for each F_{IT} value.

Source of Variation	Sums of Squares	Degrees of freedom	Mean Square	$F_{4,5}$
Between	.32663	4	.0817	41 P < 1%
Within	.00994	5	.0020	
Total	.33657	9		

9.3.3.2 Results

The F-statistics of the five populations for the *Es-3* and *Hbb* loci are given in Tables 9.9 and 9.10. Where possible, standard errors are given. The negative values of the "population" F_{ST} in the TUR6 (*Es-3*) and plague (*Hbb*) populations occurred because the sampling variance exceeded the observed variance. As there were no significant differences between the same F-statistics at the *Es-3* and *Hbb* loci where t-tests were possible, the two estimates of each F-statistic were either pooled (F_{IS}) or averaged (F_{ST} and F_{IT}) and these joint estimates are given in Table 9.11. Wright's correction did not make a great deal of difference to the F_{ST} estimates and the estimates based on Wright's correction have not been included in Table 9.11. Because Table 9.11 summarises the results of the previous two tables, this alone will be discussed.

The mean estimates of F_{ST} and F_{IT} in Table 9.11 are based on only two loci, consequently the confidence limits for each mean are wide and it is not obvious that there is significant heterogeneity across the five populations. To demonstrate heterogeneity across populations, a one-way analysis of variance can be carried out by treating the two loci (*Es-3* and *Hbb*) as replicates within each population. There are two replicates for five populations (treatments). Lewontin and Krakauer (1973) suggest that F_{ST} is distributed as a χ^2-variate and F_{IT} is a correlation. Hence, analyses of variance were carried out on transformed values of F_{ST} and F_{IT} as well as on raw values. The square-root of F_{ST} and Fishers z-transformation of F_{IT} were used. A summary of two analyses of variance are presented in Tables 9.12 ("sample" F_{ST} values) and 9.13 ("population" F_{IT} values), where it is clear that the five populations are heterogeneous. Analysis of variance tests based on raw data also showed significant heterogeneity between populations.

The two methods of calculating F-statistics sometimes give similar estimates (e.g. F_{ST} for country) but they may also yield widely divergent

values (e.g. F_{IT} for city). The most appropriate method varies between populations. For plague and country populations, the "population" F-statistics will certainly be most appropriate because the conditions of binomial sampling are usually met in these populations. For the Turretfield populations the "sample" F-statistics may be preferred because most of the population was enumerated, although perhaps a Gaussian correction of (n/(n-1)) should be applied to the F_{ST} values. The three F_{IT} estimates are very similar in the populations with roughly equal sample sizes (plague and Turretfield 6 and 7) but are less consistent when sample sizes are very heterogeneous (city and country).

Turning to a comparison between populations, it is not surprising that F_{IS} is negative in the plague and Turretfield 6 and 7 populations since these samples are almost all from single sites. F_{IS} is positive for the city and country populations because these contain samples from two or more sites.

The relative values of F_{ST} are much as predicted earlier on the basis of observations of the ecological attributes of the populations. The plague population has the smallest F_{ST}, the Turretfield populations are slightly larger and F_{ST} is very large in the city and country populations.

The large F_{IT} values for city and country populations are evidence for high levels of inbreeding due to population subdivision in these populations, but the negative values in the plague and Turretfield 6 and 7 populations are unexpected.

9.3.3.3 Variation between subpopulations at the *Agouti* locus

Because the two alleles at the *A* locus are dominant and recessive, full genotype data are not available and only F_{ST} can be estimated. However, the problem of gene frequency estimation must be overcome first. If the fixation index in a sample is zero, the frequency of the recessive allele (when the recessive homozygote occurs with frequency r_i) is simply estimated from

$$q_i = \sqrt{r_i} \qquad \ldots\ldots(9.9)$$

But if the fixation index in the sample is F_i, then the maximum likelihood estimate becomes

$$q_i = (-F_i \pm \sqrt{F_i^2 + 4r(1-F_i)})/2(1-F_i) \qquad \ldots\ldots(9.10)$$

and the estimated sampling variance of q_i is

$$\sigma_i^2 = r_i(1-r_i)/N_i(F_i + 2q_i(1-F_i))^2 \qquad \ldots\ldots(9.11)$$

From the earlier results on F_{IS} in different classes of samples, it is clear that the maximum likelihood formula should be used with the $\hat{F}_{IS}$ estimate appropriate to the relevant class of sample. However, the method is open to objection when $r_i=0$, because $q_i = (-F_i \pm F_i)/2(1-F_i)$ and if F_i is negative, this yields,

$$q_i = 2F_i/2(1-F_i) \text{ or zero}$$

and I will take the solution,

$$q_i = 2F_i/2(1-F_i) \qquad \ldots\ldots(9.12)$$

This solution will be satisfactory when q_i is being estimated over a number of samples and only the mean and variance of q_i are of interest. Unless the method given above is used, when F_i is negative q_i will be consistently underestimated and the overall mean and variance will be biassed. But the method can be very unsatisfactory because it will infer that these are (or should be) recessive genes in some samples that do not contain any recessive genes in heterozygotes.

To show how the gene frequency estimates change as F_i is assumed to be zero or not, sample gene frequencies were estimated both ways. The results are shown in Table 9.6. As might be expected from the above equation for q_i, the greatest change in estimated gene frequencies occurs when F_i is large negative and r_i is zero. With two estimates of q_i for

TABLE 9.14 Estimates of F_{ST} for *A* locus.

		Plague	TUR6	TUR7	City	Country
Sample, assuming $F_i = 0.0$		.011	.453	.246	.262	.304
Sample, $F_i \neq 0.0$		.008	.236	.132	.274	.330
Population	(a)	.007	.275	.103	.397	.393
	(b)	.007	.255	.099	.371	.368

each sample, two estimates of F_{ST} can be made for each population, and this was done for the "sample" F_{ST} method of calculation. For the "population" F_{ST} estimate the maximum likelihood estimate of q_i with $F_i \neq 0$ was used and all the results are in Table 9.13. In calculating the "population" F_{ST} estimates, the following formulae were used for N_o, s_q^2 and σ_s^2:

$$N_o = (\sum_{i=1}^{n} N_i - \sum_{i=1}^{n} N_i^2 / \sum_{i=1}^{n} N_i)/(n-1) \quad \ldots\ldots(9.13)$$

$$\bar{q} = \sum_{i=1}^{n} N_i q_i / \sum_{i=1}^{n} N_i \quad \ldots\ldots(9.14)$$

$$s_o^2 = \sum_{i=1}^{n} N_i (q_i - \bar{q})^2 / n \quad \ldots\ldots(9.15)$$

$$\sigma_s^2 = \sum_{i=1}^{n} N_i \sigma_i^2) / N_o (\sum_{i=1}^{n} N_i - 1) \quad \ldots\ldots(9.16)$$

where σ_i^2 was obtained from Equation (9.11).

From Table 9.14 can be seen that assuming $F_i=0$ leads to enlarged F_{ST} values in the Turretfield populations, but does not make much difference in the other populations, probably because F_i is smaller, on average. With the exception of the plague population, the F_{ST} estimates for the *A* locus are larger than for the *Hbb* and *Es-3* loci. Application of Lewontin and Krakauer's (1973) test for homogeneity of F_{ST} estimates to each population (Table 9.15), shows significant heterogeneity in the plague and TUR6 populations only. One problem with the Lewontin-Krakauer test is that the theoretical variance depends upon $\bar{F}_{ST}$ (the mean F_{ST}) and if $\bar{F}_{ST}$ is small, then relatively trivial changes in $\bar{F}_{ST}$ due to different methods of calculating F_{ST} can seriously change the results of the test. For the plague population, if the test is applied to "sample" F_{ST} estimates, $\bar{F}_{ST}$ is increased, s_F^2 is reduced and the test Chi-square reduces to $\chi_2^2 = 1.16$. ence the significant χ_2^2 obtained for the plague population may be an artifact of the method of calculating F_{ST}.

TABLE 9.15 Lewontin and Krakauer's test of homogeneity of F_{ST} estimates applied to "Population" F_{ST} estimates for *Hbb*, *Es-3* and *A* loci.

	Plague	TUR6	TUR7	City	Country
Hbb	-.003	.041	.031	.209	.152
Es-3	.009	-.026	.019	.146	.197
A	.007	.275	.103	.397	.393
n	6	4	4	6	6
$\bar{F}_{ST}$	.0043	.0967	.051	.251	.247
S_F^2	.0000413	.02497	.00203	.0171	.0164
$\sigma_F^2 = 2\bar{F}_{ST}^2/(n-1)$	.0000074	.00623	.00173	.0252	.0244
$\chi_2^2 = 2(S_F^2/\sigma_F^2)$	11.2**	8.01*	2.31	1.35	1.34

* Probability < 5%

** Probability < 1%

The Turretfield populations offer an interesting contrast because the χ^2_2 test is significant for TUR6 but not TUR7, suggesting that the significant χ^2_2 is the result of temporary heterogeneity.

The inclusion of F_{ST} (*A* locus) in the mean ($\bar{F}_{ST}$) values for each population (Table 9.15) inflates the $\bar{F}_{ST}$ values relative to Table 9.11. The difference between the plague population and the two Turretfield populations is accentuated and the city and country populations now appear more highly inbred with $F_{ST} \doteq 25\%$.

9.3.4 Genetic distance and geographic distance between subpopulations

Many genetic distance statistics have been proposed to measure the difference between two samples. Probably the most recent is that of Latter (1973a), denoted ϕ^*_{ij}, which has the advantages of being simple to compute and of providing a measure of divergence related to the coefficient of kinship (Latter, 1973b). For these reasons it will be used in this section to measure genetic distances between samples in the five mouse populations. ϕ^*_{ij} was calculated with the formula:

$$\phi^*_{ij} = \tfrac{1}{2}\left(\sum_{k=1}^{\ell} (q_{ik}-q_{jk})^2\right)/\left(1 - \sum_{k=1}^{\ell} q_{ik}q_{jk}\right)$$

where q_{ik} is the frequency of the kth of ℓ alleles in the ith sample. ϕ^*_{ij} was calculated between each pair of samples within each population for the five loci: *Es-3*, *Hbb*, *A*, *Es-1* and *Erp-1*; and the five estimates were averaged to yield an overall measure of genetic divergence. For the *A* locus, gene frequencies were estimated by the simple method:

$$q_{i1} = \sqrt{r_i}$$

$$q_{i2} = 1-q_{i1}$$

The geographic distance between each pair of samples within a population was measured in straight lines overland, except for Wardang Island where the distances were measured to the nearest point on the

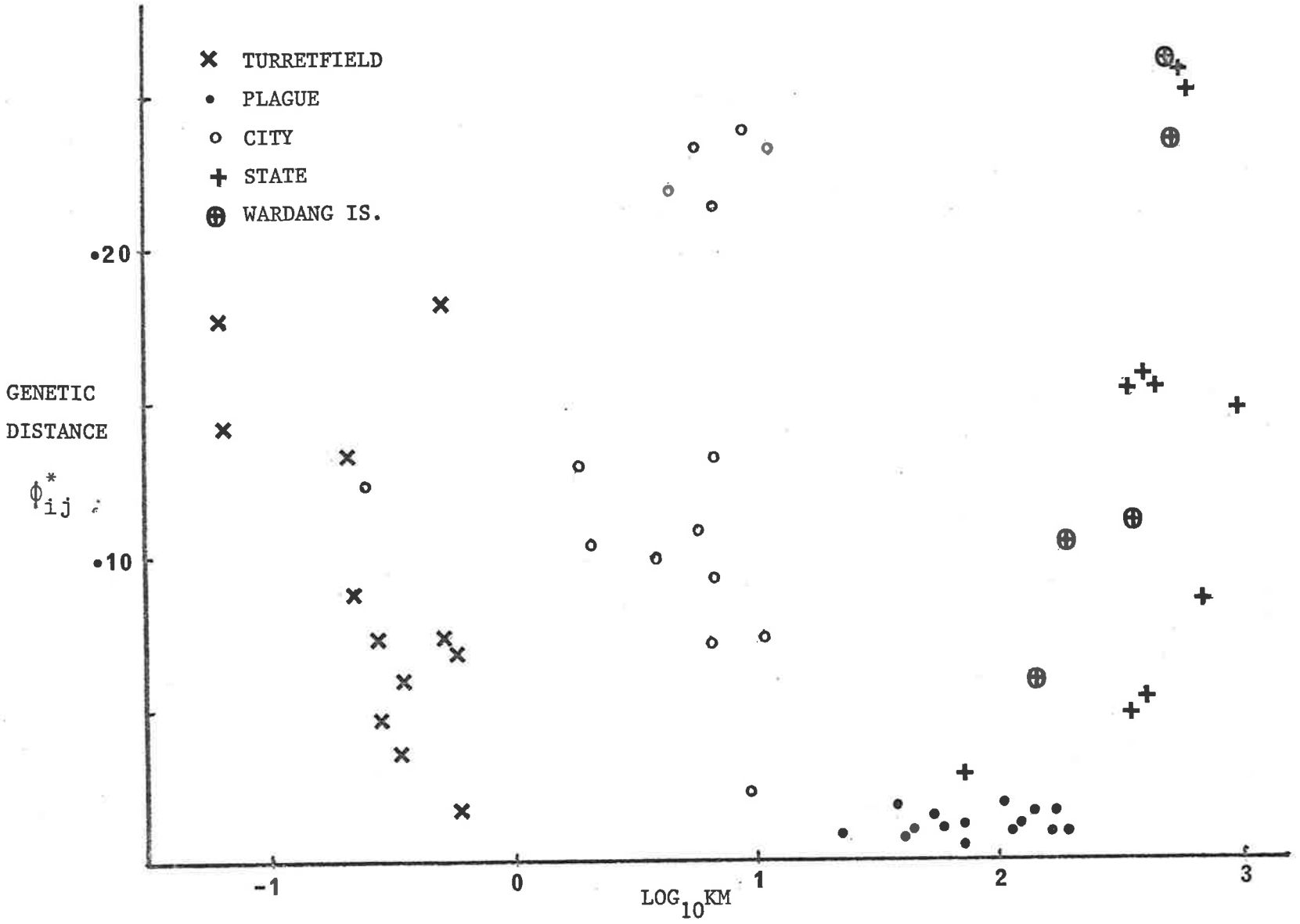


FIGURE (9.1); ϕ^*_{ij} PLOTTED AGAINST LOG_{10} KM

mainland. Because the geographic distance between subpopulations varies from 70 metres to 940 km, it is difficult to represent the regression of ϕ^*_{ij} on geographic distance with an arithmetic scale for km. Hence a logarithm scale for km is used in figure (9.1) which presents the results for the four South Australian populations.

Three comments can be made about the results shown in Figure (9.1):

(a) the plague population is very different from the other populations in that the values of ϕ^*_{ij} are small with a low variance;

(b) in none of the populations is there a significant increase in genetic distance with increasing geographic distance although significance is approached in the country population. The regression coefficients are:

plague:	b = +.00001 ,	t_{13} = .43	p>.50
country:	b = +.00018 ,	t_{13} = 2.11	p>.05
city:	b = +.002 ,	t_{13} = .34	p>.50
Turretfield 6+7	b = -.13	t_{10} = 1.56	p>.10

The points involving Wardang Island are indicated in figure 9.1 and there is no marked difference between the genetic distances involving Wardang Island and those for mainland samples only.

(c) For the Turretfield, city and country populations overall there is only a slight tendency for genetic distance to increase with geographic distance.

To find out whether or not the lack of regression between ϕ^*_{ij} and distance in km was unique to South Australian mouse populations, data from Ontario (Petras *et al.*,1969) and Texas (Selander, Yang and Hunt, 1969) were also used to calculate ϕ^*_{ij}. Geographic distances could be estimated in each case from maps that were provided to show the location of each sampling site. From Petras *et al.* (1969) gene frequency data on six loci (*A, Es-2, Es-3, Es-5, Hbb* and *Ldr*) from six farms (farms B,C,D,F,G and H provided adequate sample sizes) were used. From Selander, Yang and Hunt (1969), the gene frequency data given in Tables 4 and 5 on

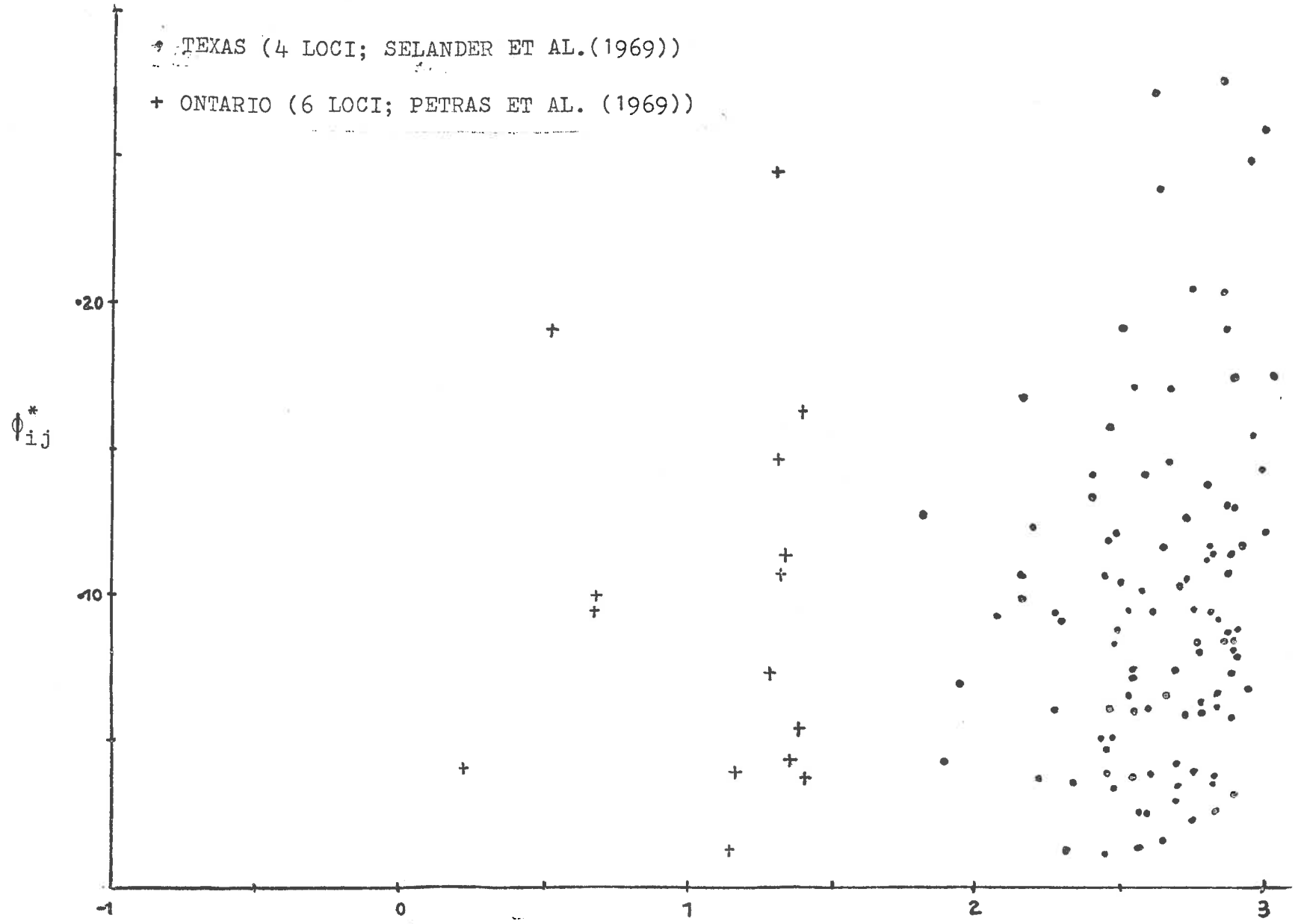


Figure (9.2); ϕ^*_{ij} PLOTTED AGAINST LOG_{10} KM

the *Es-2*, *Es-3*, *Es-5* and *Hbb* loci were taken from 15 regional locations where more than 80 mice were scored. Within each region, between 2 and 15 different sites were sampled. The mean ϕ^*_{ij} over all loci was calculated for each pair of subpopulations sampled within each population and plotted against $\log_{10}$km in Figure (9.2). The regression of ϕ^*_{ij} on km is significant for Texas, but not for Ontario. For Texas,

$$\phi^*_{ij} = .0708 + .0000606\text{km}; \quad (t_{103} = 2.41, \; p<5\%)$$

Inspection of figures (9.1) and (9.2) shows that the Ontario population is similar to the city population and that the Texas population is similar to the South Australian country population, both in values of ϕ^*_{ij} and of km. The conclusion which can be drawn from this analysis of the relationship between genetic and geographic distances in house-mouse populations, is simply that geographic distance has no detectable effect on genetic distances until a distance of about 120 km is exceeded. Under that distance (in the TUR, city, Ontario and plague populations) the magnitude of ϕ^*_{ij} depends upon gene flow and population structure (compare city and plague) and also sample sizes (small samples will inflate ϕ^*_{ij}; this fact may explain the rather large values of ϕ^*_{ij} in the TUR6 and TUR7 populations). In populations where geographic distances exceed 120 km (country and Texas), genetic distance does increase with geographic distance.

9.4 DISCUSSION

This study of the genetic structure of four different mouse populations has shown that the genetic structure of a population will change when the ecological conditions affecting population density and dispersal rates are altered. The differences between the 1970 plague and the 1972 Turretfield populations show that even in the same area, population structure changes with time (this is better illustrated in Chapter 10). The great differences in F-statistics between the plague and the city populations emphasise the role that ecological conditions play in determi-

ning the genetical structure of populations. In the first case there is a population living in apparently optimal conditions over hundreds of square km of countryside for two or more generations while in the second case there is a population that ekes out an existance in a predominantly adverse environment.

These results show that it is probably invalid to make predictions about the genetical structure of other mouse populations without carrying out intensive ecological studies. Extrapolation from one mouse population to another should only be done with great care. Discussions of mouse ecological genetics which utilise data collected from different populations (including data from confined laboratory colonies) but which do not take adequate account of the ecological and genetical differences between populations, run the risk of drastically oversimplifying the real conditions. Attempts to describe a "typical" mouse population (e.g. DeFries and McClearn, 1972; Selander, 1970) could well lead to incorrect conclusions. Each mouse population should be studied as a separate entity in its own right and out of many such studies there may grow an appreciation of the potential that mice have to modify their behaviour, ecology and population genetics to suit the prevailing circumstances.

The demonstration that F_{IS} is greatly affected by the number of panmictic subpopulations (or in practical terms, sites) sampled is a validation of the mathematical theory in Chapter 3. Also, it shows that the estimated F_{IS} in a population will depend upon the care with which samples are collected from subpopulations. The negative F_{IS} values seen in the Turretfield 1972 populations and the positive values in the city and country populations exemplify this point.

On the basis of ecological evidence, the prediction was made that Nm was large in the plague population, very small in the city and country populations and intermediate in the Turretfield populations. The mean

estimates of F_{ST} in Table 9.15 are satisfactorily in inverse proportion to the estimates of Nm. F_{ST} is negligible in the plague population, intermediate in the Turretfield populations and very large in the city and country populations. This result is in accord with theoretical predictions. The effective size of subpopulations and the amount of gene flow between subpopulations play an important part in determining the amount of variation between subpopulations within a population. But I should emphasise that in the absence of precise data about N and m in the populations that I have studied, I cannot ignore the possibility the selection also played a part in determining the amount of variation between subpopulations.

The most surprising result was the failure of the prediction about the relationship between genetic and geographic distances. Theory predicts that the genetic distance should rise rapidly with increasing geographic distance and flatten out at greater distances. The observed relationship was almost the opposite: genetic distance appeared to be independent of geographic distance until about 120 km, and only then did genetic distances increase with geographic distances.

This observation implies that the island model of population structure may be quite adequate to analyse the genetical structure of mouse populations covering a region 100 km or so across. The other models of population structure predict a relationship between genetic distance and geographic distance that is not observed.

Providing an explanation for the discrepancy between observation and theory is not easy in this case. Had the results from Ontario and Texas not supported the conclusion drawn from the South Australian population, I may have suggested that the discrepancy was an artefact due to the way in which the samples were collected differently in the four populations, just as the F_{IS} values differed between populations.

But the similarity with the results from Ontario and Texas, where different loci, populations and sampling techniques were used, is proof that the observed relationship is real and not an artefact.

A more likely explanation is that differentiation of mouse subpopulations is prevented by large amounts of "long range" gene flow. As was pointed out in Chapter 6, nothing is known about the long distance movements of dispersing mice. Active movement by walking probably does not exceed a one or two km usually, but long distance movements of up to hundreds of km by passive transport (in the sense of Dobzhansky, 1973) in trucks, trains and boats carrying hay-bales, grain, etc. are possible. The numbers of mice transported passively and their reproductive success on arrival at a destination can only be guessed at present. The territorial behaviour of mice observed in captivity might well prevent emigrants from entering dense mouse populations, but an emigrant might be able to find a vacant niche where it can become established and find a mate if population densities are low. A rough calculation can show the number of immigrants needed to maintain the F_{ST} values observed. For city and country populations, $F_{ST} = 25\%$ and if $F_{ST} = 1/(1+4Nm)$ is a reasonable approximation to the real situation, Nm need only be a little less than unity to explain the observed F_{ST} values. The frequency of effective long range migration between geographically distant subpopulations need only be equivalent to 1/N, i.e. one reproductively successful migrant per generation per subpopulation. If the generation time for the average mouse population is about six months (see Chapter 10 for evidence of this at Turretfield), then two successful immigrants per year will be enough to give the observed variation between subpopulations. For city subpopulations such as the Zoo, the University, the Waite Institute and the Pet Shop, where hay, grain and other foodstuffs are used continuously and imported from the country and city warehouses, it is certainly conceiv-

able that a couple of new mice could be imported in each subpopulation each year.

Stabilising selection acting over an area within which the environment was fairly uniform, might provide another explanation. The work of Hunt and Selander (1973) provided evidence of a stable hybrid zone between the two subspecies of house-mice in Denmark. Their interpretation of this situation was that each subspecies had a coadapted gene complex which broke down when hybridisation occurred. The introgression of genes across the hybrid zone may be prevented by the relative weakness of hybrids. This observation suggests that some, at least, of the gene frequencies in a mouse population are subject to selection favouring an optimum combination of genotypes and that marked changes in gene frequency will be opposed by selection. Gene frequencies could be stabilised around certain values (Wright's selective peaks, e.g. Wright, 1970) by selection over wide geographic areas although differentiation of gene frequencies could occur locally by random sampling drift.

The selection hypothesis might also explain why genetic distances began to increase after greater geographic distances. Presumably the external environment (e.g. temperature, humidity, daylength, food sources) is gradually changing across the landscape and there may be a similar change in gene frequency optima. Hence at longer geographic distances, genetic distances may increase because of the gradual change in optimum gene frequencies due to gradual environmental changes.

I am not satisfied that an adequate explanation can be put forward for the observed relationship between genetic and geographic distances. More data from natural populations and more work on mathematical models of population structure, gene flow and selection are needed.

CHAPTER 10

A STUDY OF NEIGHBOURING MOUSE SUBPOPULATIONS AT TURRETFIELD

10.1 INTRODUCTION

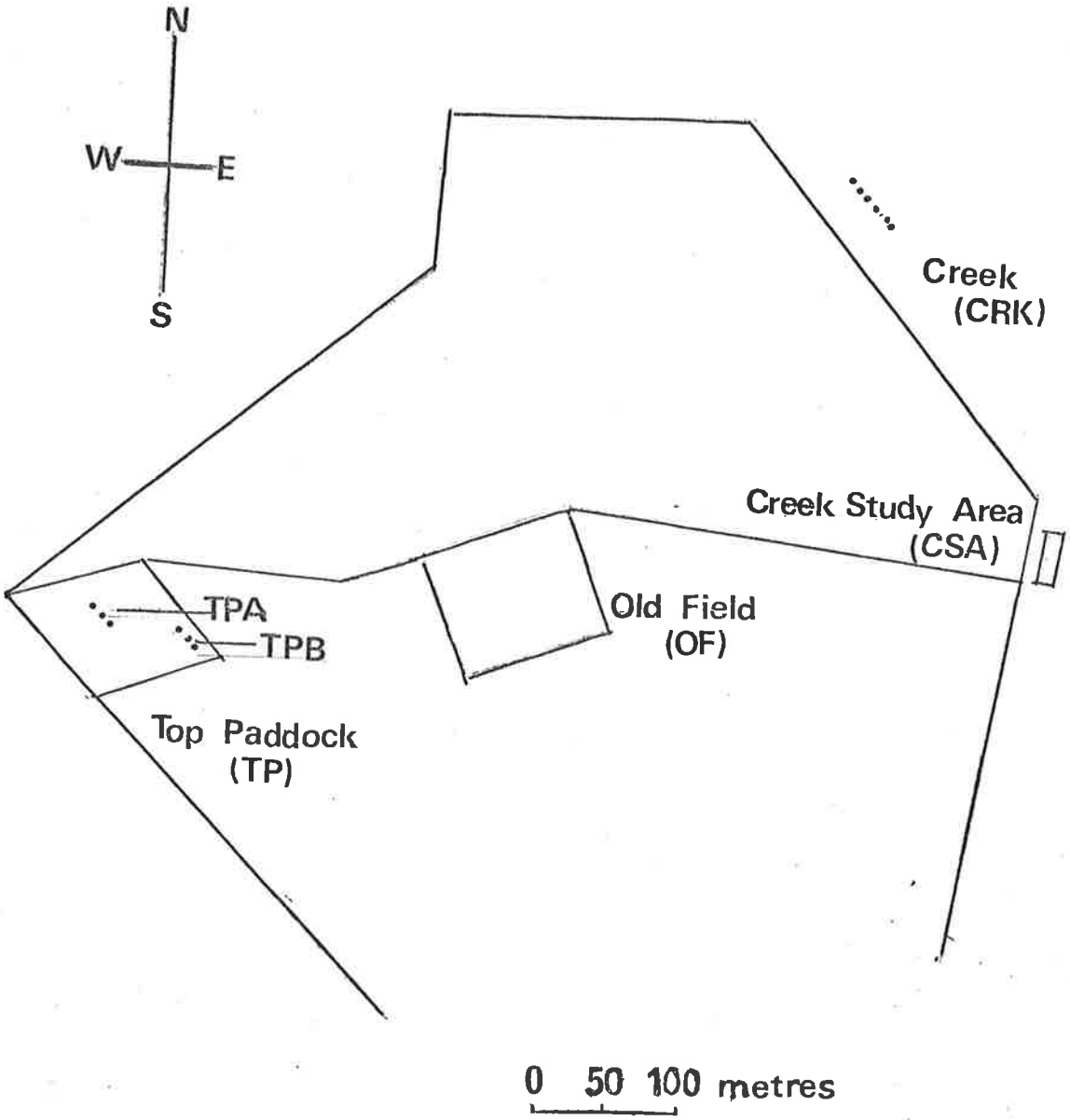
In the summer of 1971, very few mice could be caught in the farmlands north of Adelaide or on Yorke Peninsula . After the mouse plague of the previous winter (1970) the mouse populations appeared to crash severely. At Turretfield in April 1971 only five mice were caught in 262 trap-nights in the study areas used by Dr. A.E. Newsome (Newsome 1969a, 1969b). However, thriving mouse colonies were found under the feedstations established by Dr. A.E. Newsome. Each feedstation consisted of a tray covered by half a 44 gallon drum (Newsome, 1970) and grain (wheat or barley) was added to the trays at regular intervals to maintain a continuous food supply. When the present study began the mice had already dug numerous burrows under each feedstation, within which they found shelter.

The feedstations were in two habitats. (See Map 10.1 for locations of study areas). Six in Salt Creek (abbreviated to CRK) have been described in Newsome (1970). Another six were set in a paddock (top paddock or TP) 600 metres S.W. of CRK and were arranged in two groups of three, 68 metres apart. In both TP and CRK the distance between neighbouring feedstations was about 10 metres (range 7.3 to 14).

Because there were few mice in the surrounding fields the feedstations contained relatively isolated mouse colonies which were suitable for a study of population subdivision into small subpopulations. The two groups of feedstations in TP were found to be separate colonies of mice (denoted TPA and TPB) and CRK was a single colony. It was feasible to obtain virtually all of the mice in a colony by trapping inside the feedstations. As surplus food and permanent shelter eliminated two

MAP 10.1

Study Area at Turretfield



of the major environmental variables affecting mouse population dynamics (Newsome, 1969a, 1969b, 1970) it was expected that a stable population structure might be established.

The intention of this project was to collect ecological and genetical data which could lead to independent estimates of the effective population size and migration rates for these mouse colonies.

10.2 ECOLOGICAL STUDIES

10.2.1 Methods

The initial plan was to make a census every six weeks. But in summer when the populations appeared ready to increase numerically, the time between censuses was shortened to four or five weeks and at other times adverse weather (very hot or cold and wet weather led to deaths in traps) made trapping unwise and the interval lengthened to seven, eight or ten weeks. In each feedstation, depending upon the number of mice thought to be present, four to six Longworth traps were set on the wire mesh protecting the grain in the tray and one to three traps were set outside within one metre of the cover. Very rarely were all traps at a feedstation entered in a single night so it was assumed that all mice had access to traps every night. In the CRK area additional traps were set between the feedstations and alongside six posts on the western bank of this study area (Newsome, 1970) to check for movements between feedstations and for mice living independently of the feedstations. In the TP area, 30 posts were arranged in a grid pattern that surrounded the two groups of feedstations and extended over part of the field distant from the feedstations. These were used as trapsites to check for mice living in the field or recently evicted from the feedstations. Except when weather was wet and the ground flooded, 15 to 29 traps were set in this area.

For control populations, two other areas were trapped: Newsome's creek study area (CSA) (Newsome 1969b, 1970) and a 1.13 hectare area in the paddock near TP that was the site of Newsome's earlier work (Newsome, 1969a). This site, which was fenced off and has remained undisturbed since 1967, resembles an old-field habitat and so became known as the old-field (OF) site. These two sites were usually lightly trapped (10 to 21 traps).

Trapping normally extended over four nights for each census, although adverse weather caused termination after three nights in September, 1971, February and July, 1972.

Following Newsome (1969a), the length of a stretched out mouse from the tip of the nose to the base of the tail (HB length) was used as an index of age. The reproductive condition of every mouse caught was assessed. Males were classified by the location of the testes (abdominal, scrotal or uncertain) (De Long 1967, Newsome 1969a,1969b) and females were examined to see whether or not the vagina was open, the nipples were prominent and hairless (indicating lactation) or the belly swollen as in pregnancy.

Each mouse trapped for the first time was lightly anaesthetised with ether, bled from the sub-orbital sinus of the right eye, checked for sexual condition, measured for HB length on a millimetre ruler and toe-clipped for future identification. When recaught, each mouse was checked for breeding condition and measured again.

In March, 1972, an intensive study of the OF site was carried out to assess the population density and degree of subdivision of the colony into family units. Traps were arranged in a 12x12 grid with a spacing of 8.22m in the N.W.-S.E. axis and 9.14m in the S.W.-N.E. axis.

The study lasted 50 weeks from July 1971 to July 1972.

FIGURE 10.1

Growth of female mice at Turretfield

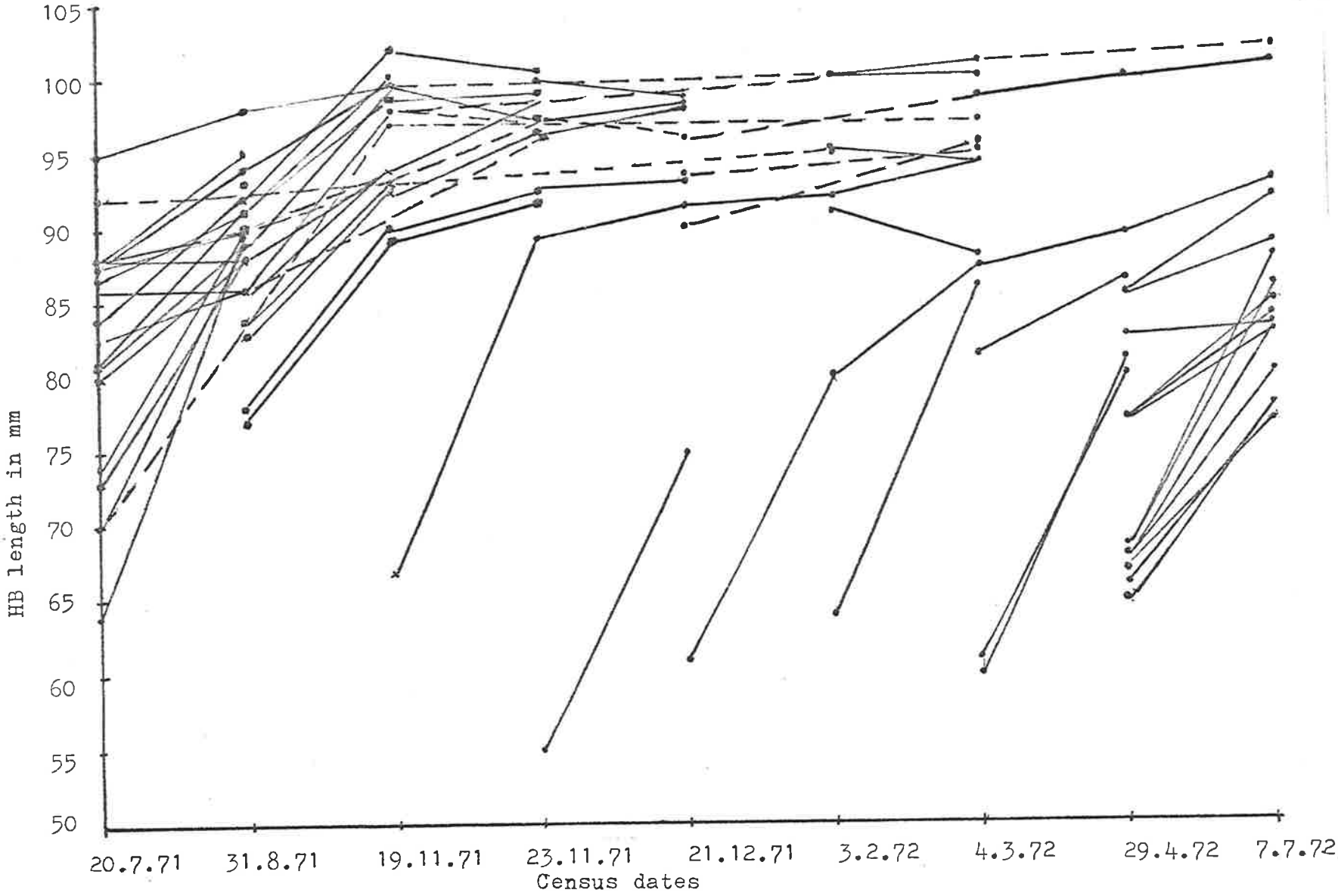
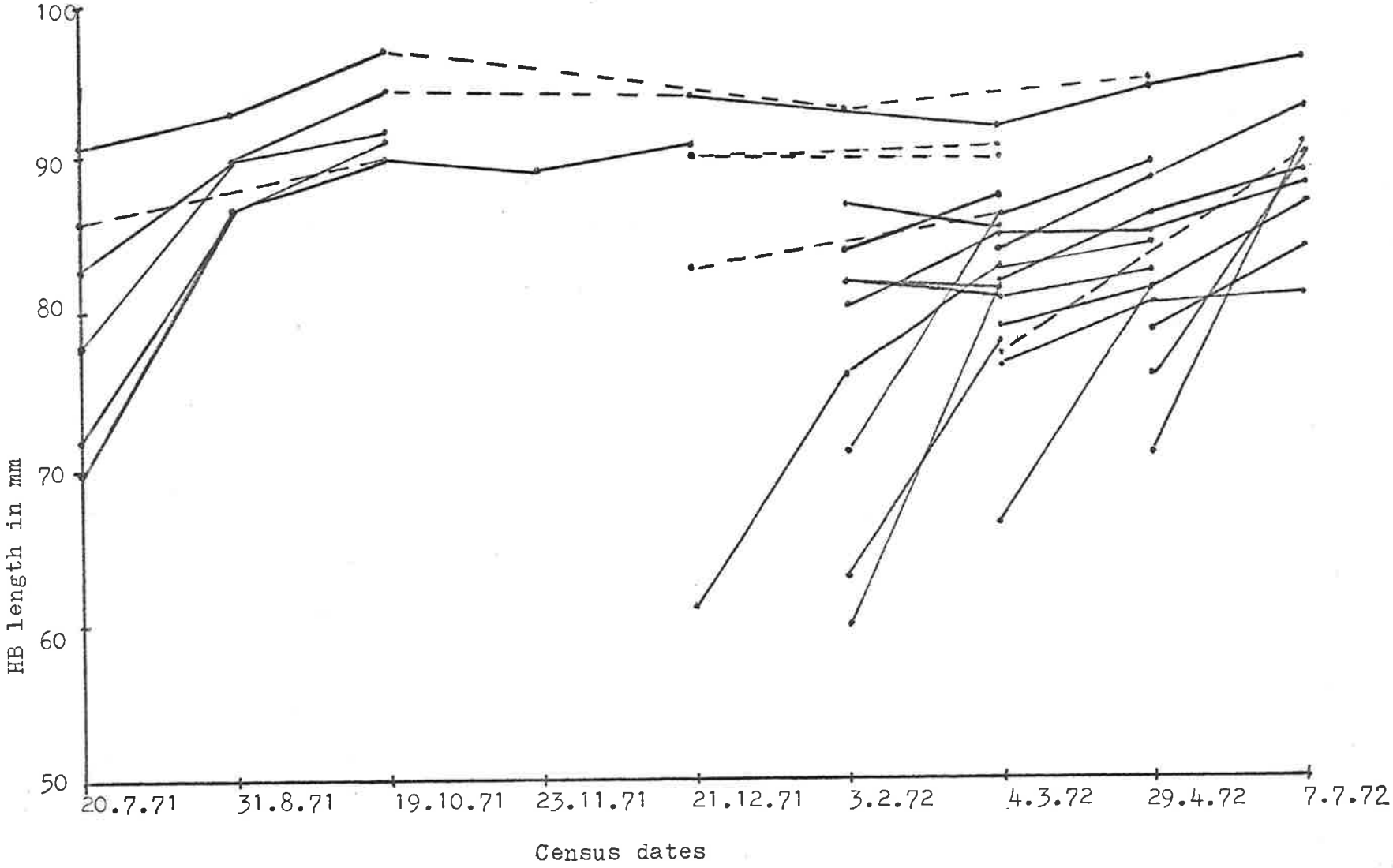


FIGURE 10.2
Growth of male mice at Turretfield



10.2.2 Growth and age of mice.

The accuracy of a single HB length measurement can be assessed from HB measurements made on mice caught on successive days during a single census. Assuming that growth is negligible, these measurements represent estimates of the same length with variation around the mean due to errors of measurement. On 34 occasions a mouse was measured on four consecutive days and an analysis of variance on these data gave an error mean square of 1.89mm, so the 95% confidence limits for a single measurement are $\pm$ 2.73mm.

The HB lengths for all mice caught on two or more censuses are presented in Figures 10.1 and 10.2 for females and males respectively. Growth of small mice is rapid, but in older mice growth almost stops and measurement errors lead to fluctuations in HB length.

If ℓ_x represents the HB length of a mouse at census in the x^{th} week of the study period and ℓ_{x+n} is the length at the next census, then the change in length ($\Delta\ell_x$) is ($\ell_{x+n} - \ell_x$) and the time interval (or *epoch*) over which the change is measured is n weeks. When $\Delta\ell_x$ is plotted against ℓ_x, an approximately linear regression is obtained and this can be used to obtain a growth equation as follows. Let the regression be $\Delta\ell_x = a - b\ell_x$ where a and b are fitted constants and measurements have been made over epochs of n weeks. Then the difference equation $\ell_{x+n} = a + (1-b)\ell_x$ can be solved by standard methods (e.g. Maynard-Smith, 1968) to yield a general solution

$$\ell_t = A + B(1-b)^t$$

where ℓ_t is the length of a mouse at a time t epochs after an initial length ℓ_o. For the situations of biological interest, $0<b<1$ and hence ℓ_t tends to A as t tends to infinity. Growth stops ($\Delta\ell_x=0$) when $\ell_x = a/b$ and this is the value that will be taken by A. We may take ℓ_o as the smallest size at which mice become easily trappable, which

is about 60mm. This is approximately the length of a three week old mouse (from Newsome's (1969a) growth curve and personal observation).

Hence $\ell_t = a/b + (60 - a/b)(1-b)^t$ is the growth curve.

This equation is based on the time interval over which the measurements were made. Data for the regression can be obtained by repeatedly measuring individual mice with a constant time interval, or by measuring a number of mice of different sizes over a single time interval or by a combination of these two methods, as in my study. Unfortunately, my data may not fulfill the basic assumptions of constant growth conditions and constant time interval because growth rates may vary with the seasons (De Long, 1967) and the time interval between measurements was not kept constant. Visual examination of the regression data suggests that growth rates were similar within each sex at all times except for females in the epoch May to July, 1972.

When growth data were collected over an epoch different from six weeks, the observed $\Delta\ell_x$ values were corrected to a six week epoch by multiplying by $\frac{6}{y}$, where y was the actual epoch in weeks. This will bias the estimate of $\Delta\ell_x$ up or down depending on whether y is less than or greater than six, respectively. An estimate of the magnitude of this bias was obtained from the regression for females over the May to July 1972 period, when the epoch was 10 weeks. The growth curve was calculated for $\Delta\ell_x$ values based on a 10 week epoch and repeated after they were corrected to a six week epoch. At an extrapolated age of 23 weeks, the two growth curves differed by less than 2½%, which suggests that the bias is relatively unimportant compared with the measurement error.

Because the growth curves are of use only for smaller mice and older mice grow irregularly, data on mice longer than 90mm were ignored in calculating regressions.

FIGURE 10.3
Regression of ΔL on L for female mice

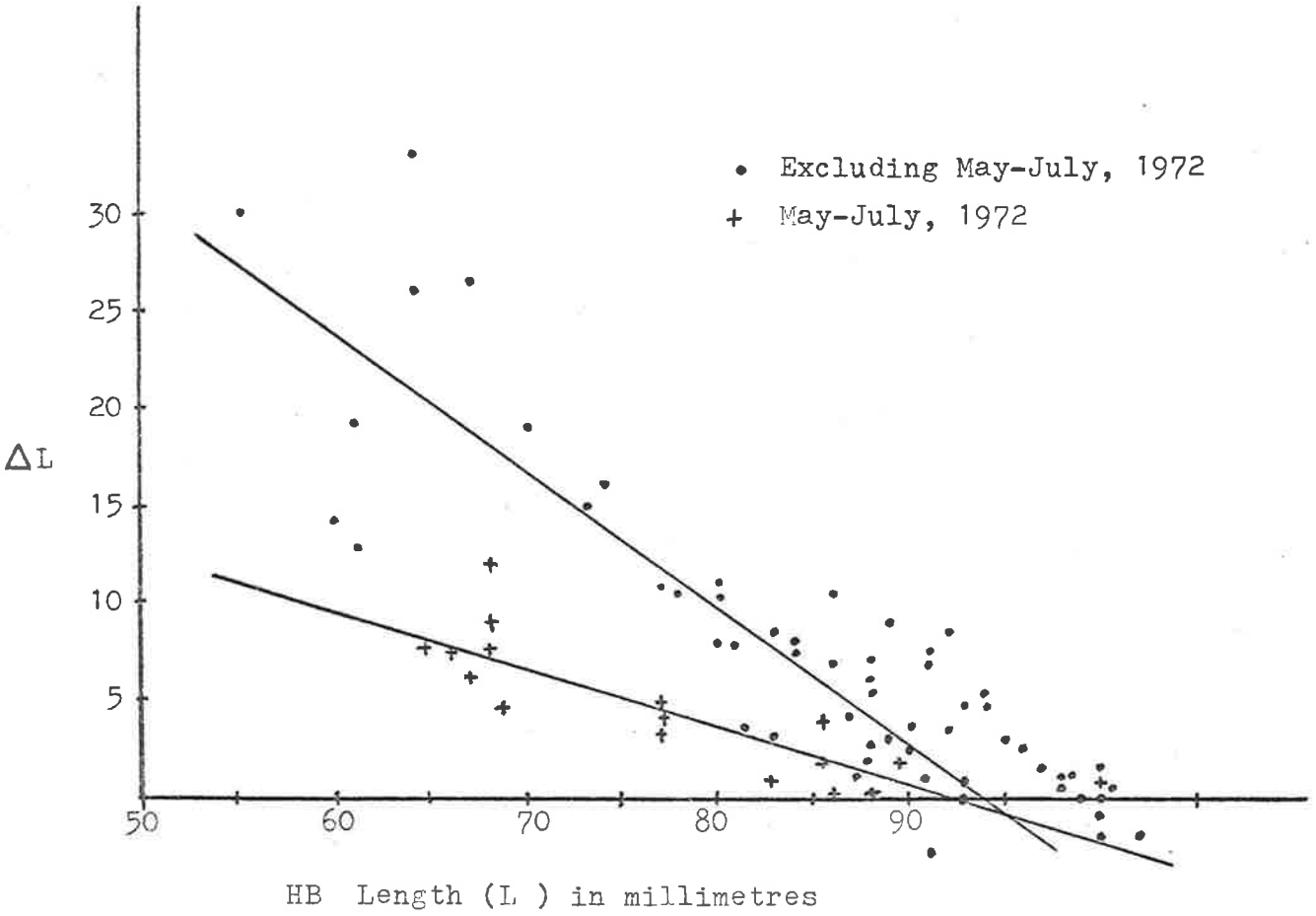


FIGURE 10.4

Regression of ΔL on L for male mice

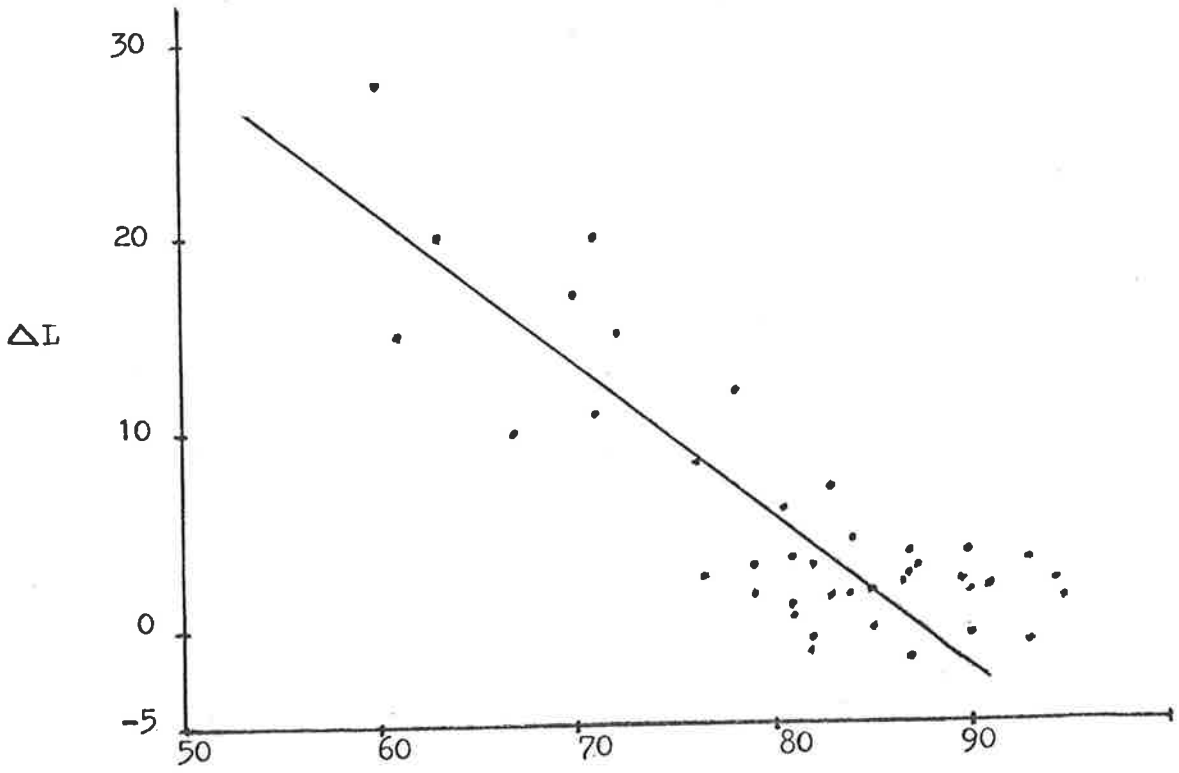
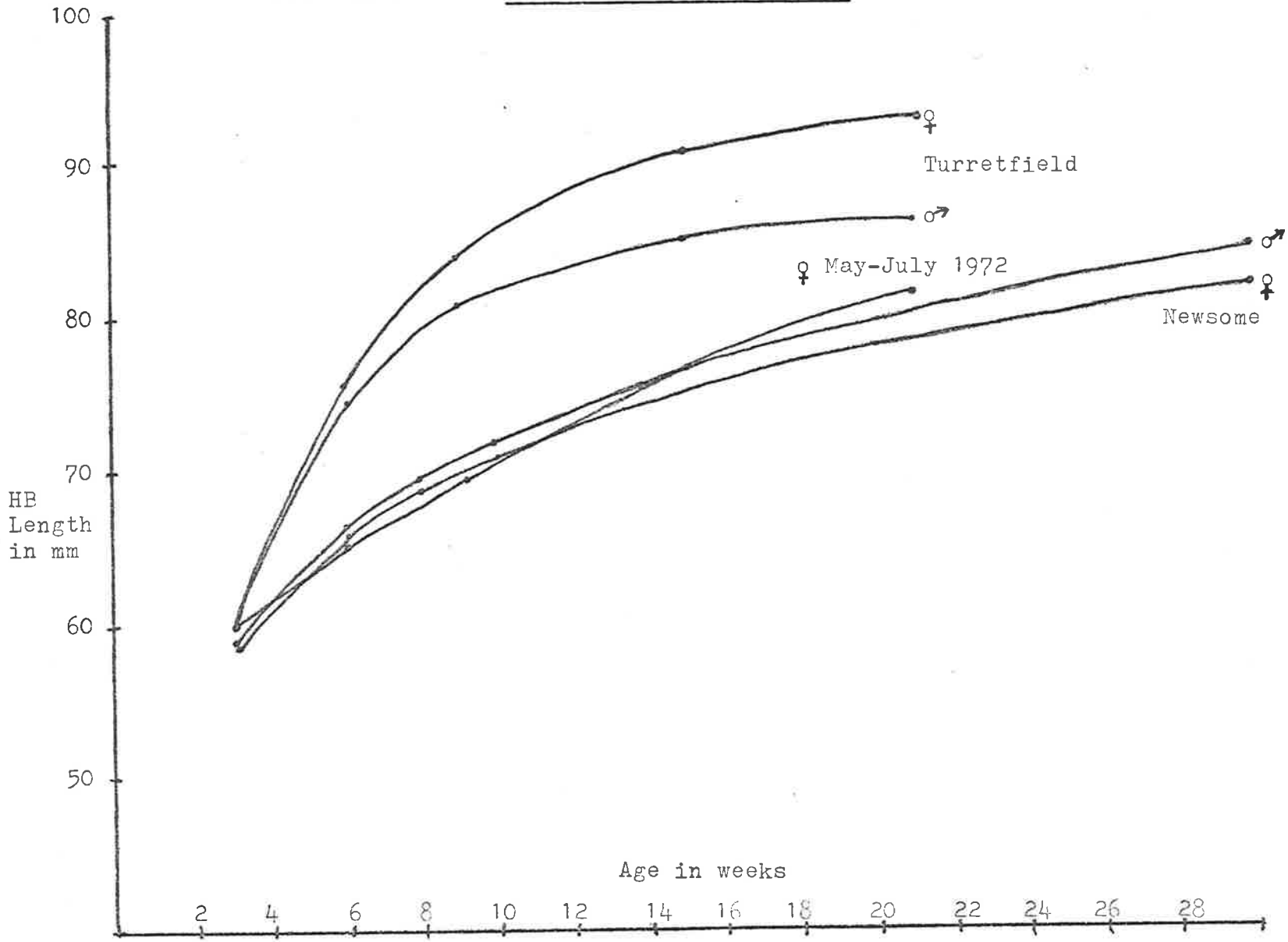


FIGURE 10.5 GROWTH CURVES FOR WILD MICE



The following regressions of $\Delta\ell_x$ on ℓ_x were calculated (see Figures 10.3 and 10.4):

(i) Females, excluding the May to July 1972 period

$$\Delta\ell_x = 66.335 - 0.709\ \ell_x$$

(ii) Females for the May to July, 1972, period

$$\Delta\ell_x = 27.214 - 0.294\ \ell_x$$

(iii) Males

$$\Delta\ell_x = 68.581 - 0.789\ \ell_x$$

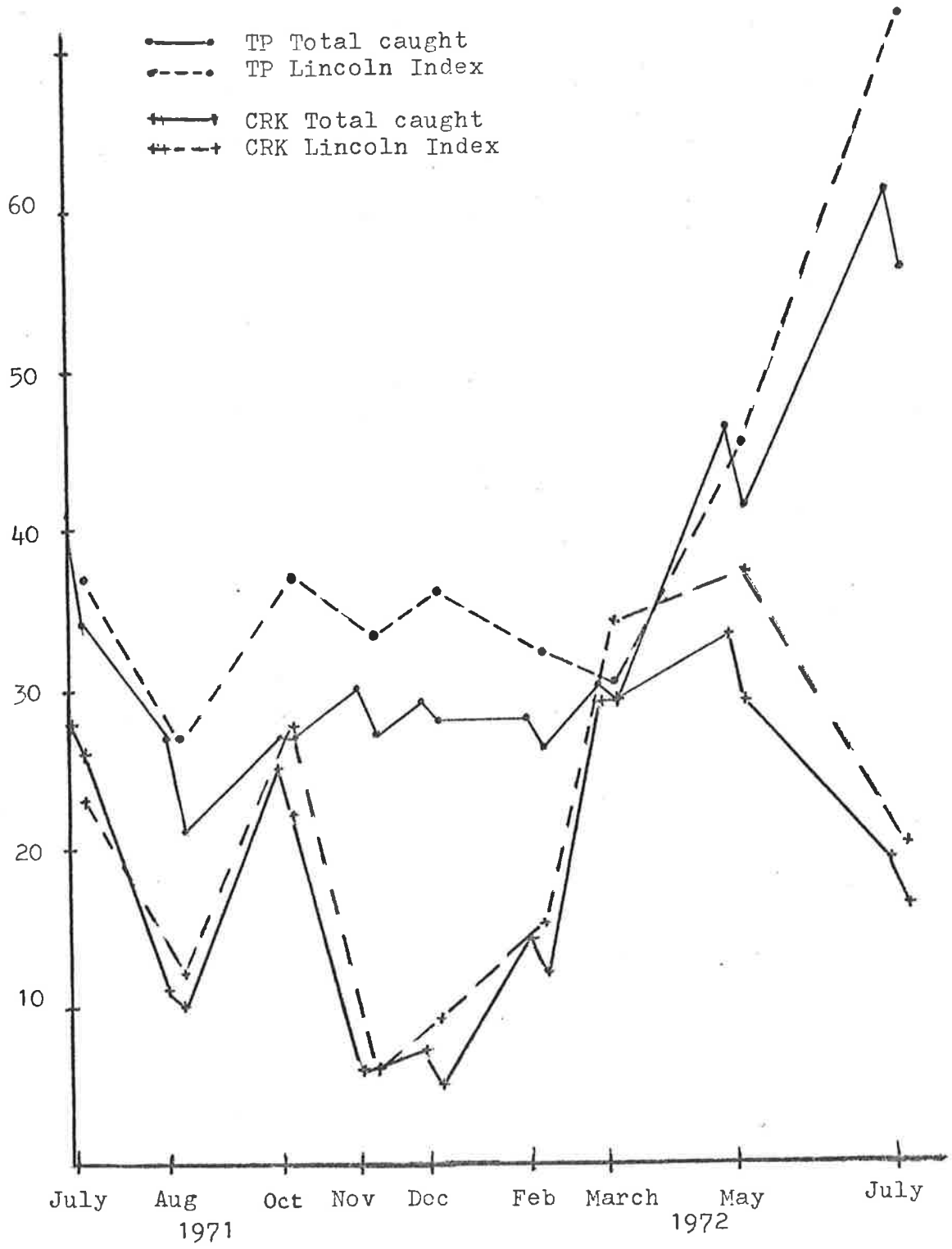
All of the regression coefficients were highly significant and the two female regression coefficients were significantly different. The slope of the regression line for males did not differ significantly from that for the female (i), but there was a highly significant difference in α values.

The growth curves derived from the above regressions (with t=0 taken as three weeks of age) are given in Figure 10.5 alongside the growth curves from Newsome (1969a). The growth rate of female mice in winter 1972 is very similar to that of Newsome's penned mice but for most of the year mice maintained under the feedstations grow faster than mice in crowded pens. De Long (1967) observed that growth rates of feral mice declined in the autumn-winter period but this was less pronounced when he alleviated the food shortage that normally occurs at this time.

One distinctive feature of the mice in feedstations is the smaller size of males. In Newsome's mice and laboratory-reared mice with wild-caught ancestors the males are usually as large or larger than the females. But Berry (1970) found that in many wild mouse populations in Britain, females were larger than males.

FIGURE 10.6

Estimated population size in TP and CRK



10.2.3 The number of trappable mice in each colony

Trapping all of the mice in an area is difficult. Crowcroft (1966) describes how wild mice act towards Longworth traps: some enter quickly but others investigate the outside of a trap at great length and rarely enter. With births, deaths, emigration and immigration going on, it is difficult to accurately estimate the size of a population although estimators are available (Jolly, 1965, Seber, 1973). As variability in trap response occurs in mouse populations (Crowcroft, 1966), any population estimate based on trapping may be biased so there is little virtue in extensive calculations and only a simple Lincoln index (Lincoln, 1930) was used to estimate the total number of trappable mice. If m mice are marked and released alive and subsequently a sample of n mice is collected of which r are marked, the Lincoln index estimate of the size of the trappable population is

$$N = m\,n/r$$

The Lincoln index was calculated from the results of the last night of trapping at each census in TP and CRK. The total number of mice trapped in four nights and thought to be still alive was 17% less than N on average. Known mortality (due to over-anaesthetisation, death in hot or wet traps and some deaths for no apparent reason) averaged 9% per census.

Top Paddock (TP) These six feedstations are really two colonies but as they were ecologically identical they have been pooled. The total number of mice trapped at feedstations, that number minus the known deaths and the Lincoln index for the last night of trapping are given in Figure 10.6. The population declined for the census of 30.8.71 - 2.9.71 because the drums protecting the food trays had been pushed off four of the feedstations by sheep grazing in the field and the trays were flooded by rain. After this the feedstation tops were firmly

anchored with bricks at the end of each census. Also, there was heavy rain in late August and this probably flooded some burrows. Under one food tray there was a litter of eight dead mice. The population quickly recovered and stabilised at nearly 30 trappable mice until March, 1972. There was no evidence of mice colonising the surrounding field during this period as only three mice were caught away from the feed-stations in November, December and February. However, in March, 10 mice were caught in the field suggesting that conditions there were more favourable for survival. But in early May only three mice were caught in the field although many mouse holes were prominent in cracks in the soil in the vicinity of the feedstations. There was little food or cover apparent in the field at this time as sheep had grazed down to the soil in most parts of the field. It is possible that some mice were living in the field and feeding at the feedstations. By the last census the population had increased markedly and mouse holes were visible in cracks all over the field. Out of nine mice caught in the field at least three seemed to be resident there as they were recaught at least once at the same trapsite. This census was abandoned because of rain. A fortnight later I counted the number of mouse holes within three metres of each feedstation. The average number of mouse holes per feedstation was ten. Although no counts were made, it is believed that for most of the study period the number of holes would have been only one or two per feedstation. The fact that the total population did not expand until numerous holes were dug around the feedstations, is evidence that the increase in mouse numbers was due to an increase in living space in the form of additional burrows dug after the soil cracked in the autumn (cf. Newsome, 1969a, 1969b). From October 1971 to March 1972 the Lincoln index remained stable at 30 to

to 36 mice, an average of five to six per feedstation. Since there was a surplus of food, but only a restricted amount of shelter in the burrows under each feedstation, this suggests that social factors interacted with the density of mice in a small area (i.e. under a feedstation) to regulate the population size. (Lidicker 1962, Anderson 1970, Newsome 1969a, 1969b, Berry, 1970.)

Creek (CRK) The data on the number of mice trapped, survivors and Lincoln index estimates are given in Figure 10.6. The population was smaller than that of TP initially and was reduced further by heavy rains in late August 1971 which led to severe flooding. In two feedstations the food trays were flooded and under the others, varying amounts of mud were carried in by floodwaters. One feedstation no longer supported any mice and two had only one mouse each. At the next census the population had recovered but 11 out of 25 mice were caught only on the west bank which was evidently a good site for burrows, as reported by Newsome (1970). However, five weeks later the population had crashed again, with the west bank and two of the feedstations lacking any mice. A brown coloured snake (probably genus *Pseudonaja*) was found under one feedstation and a young rat was trapped under another. In the absence of any signs of epidemic diseases as noted by De Long (1967) or of further flooding, it is surmised that this population decline was due to predation. In December 1971, the colony consisted of one female under one feedstation and six juvenile mice under another. By February, 1972, the colony was recovering and one or more mice were caught under each feedstation. In March and May 1972 the numbers reached the levels of TP in summer and since few mice were trapped outside the feedstations (five in March, two in May) it appears that this colony was also stabilising at 30-35 trappable mice. Finally, in July 1972, the population declined again, due to the activities of rats

FIGURE 10.7

Trap rate in Old Field and Creek Study Area

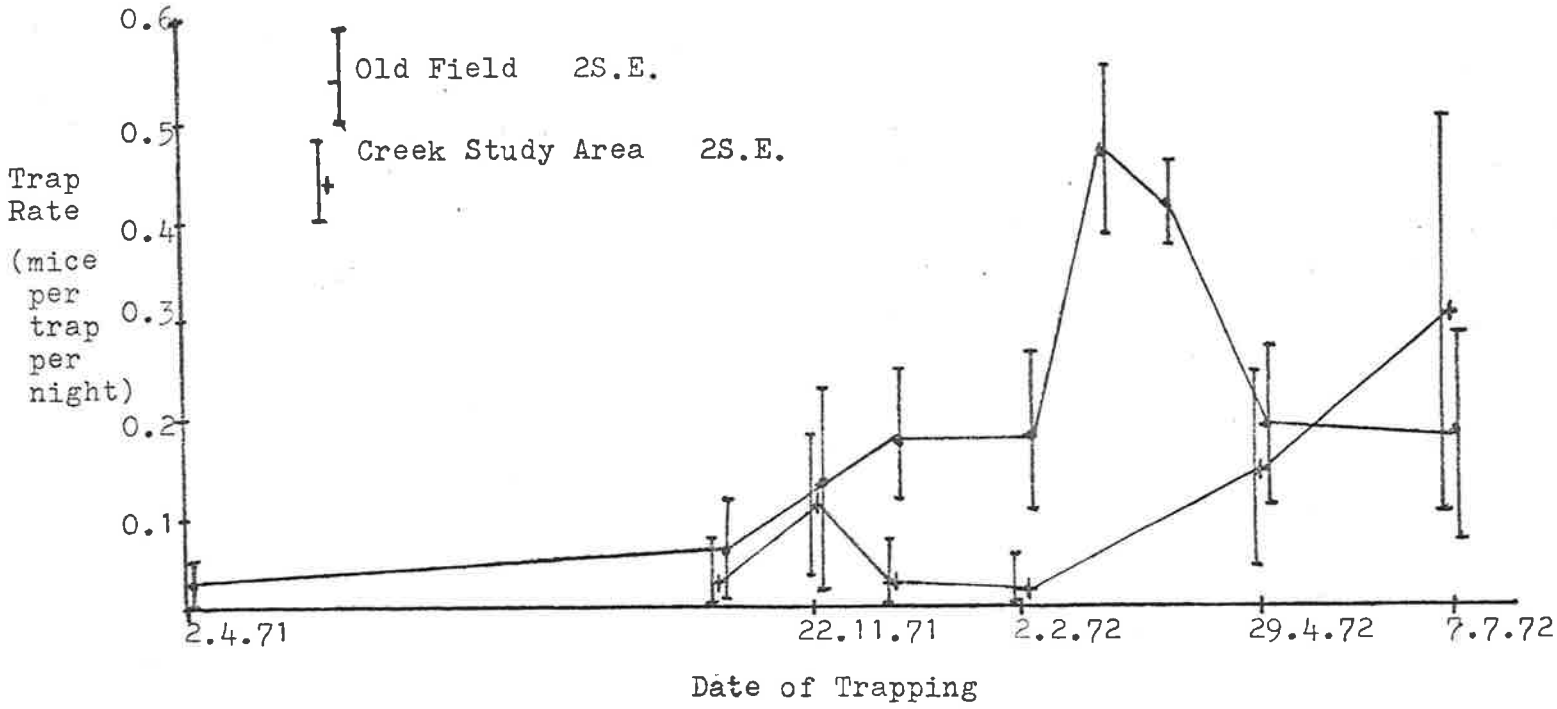


FIGURE 10.8a
Proportion of males with scrotal testes

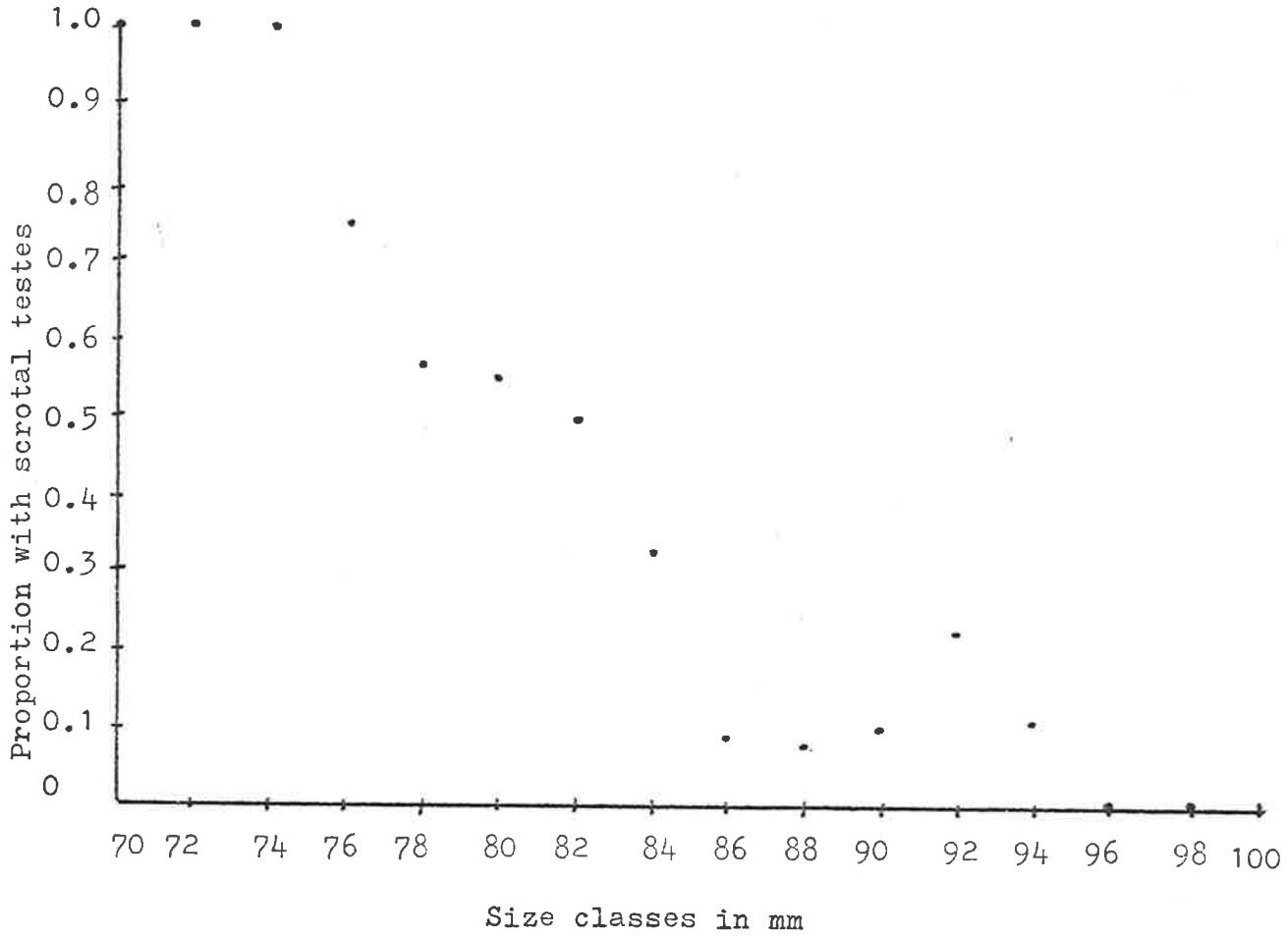
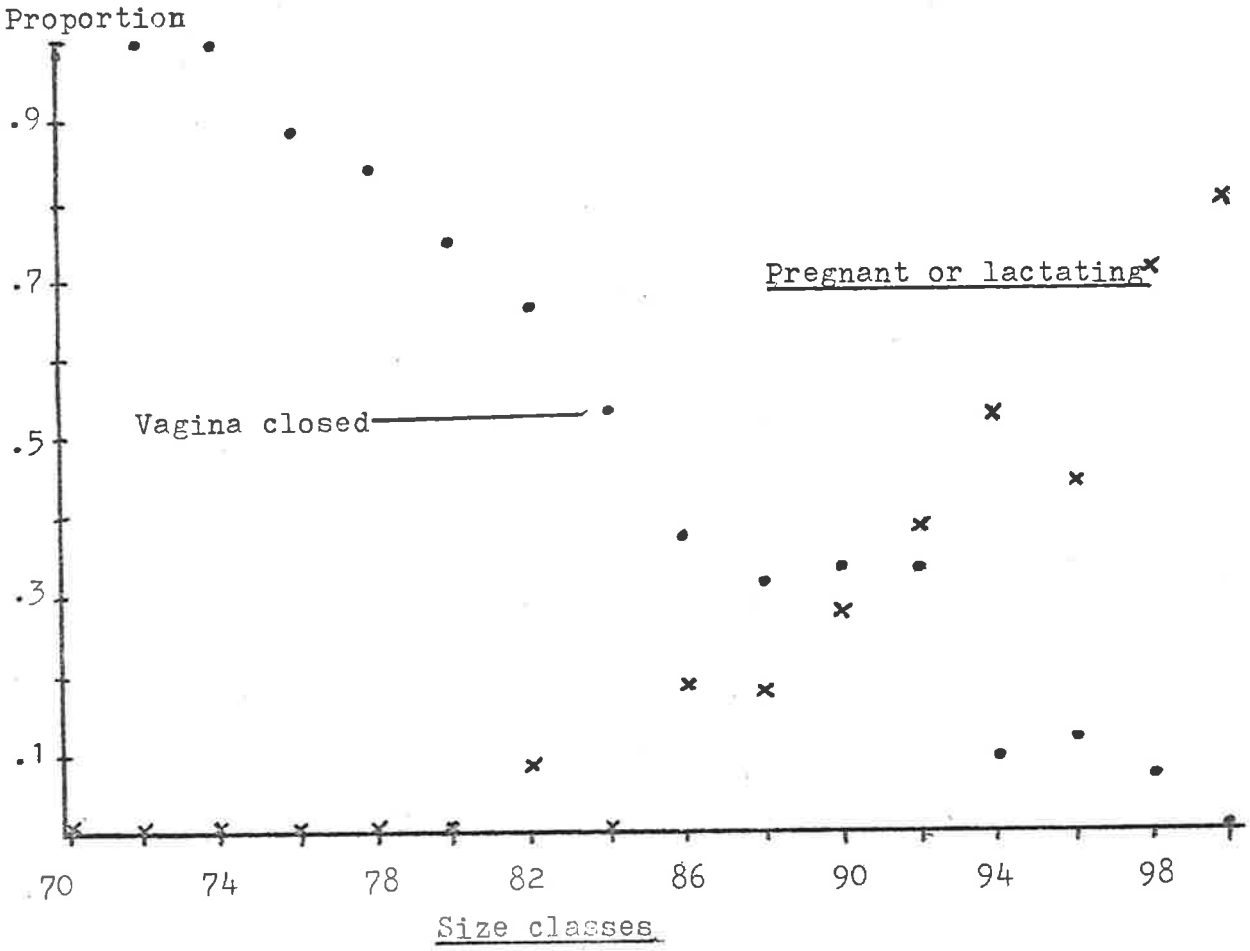


FIGURE 10.8b
FEMALE REPRODUCTIVE STATE



that usurped the mice and dug large burrows under three feedstations. One rat was caught and identified as a Black rat *Rattus rattus*, L.

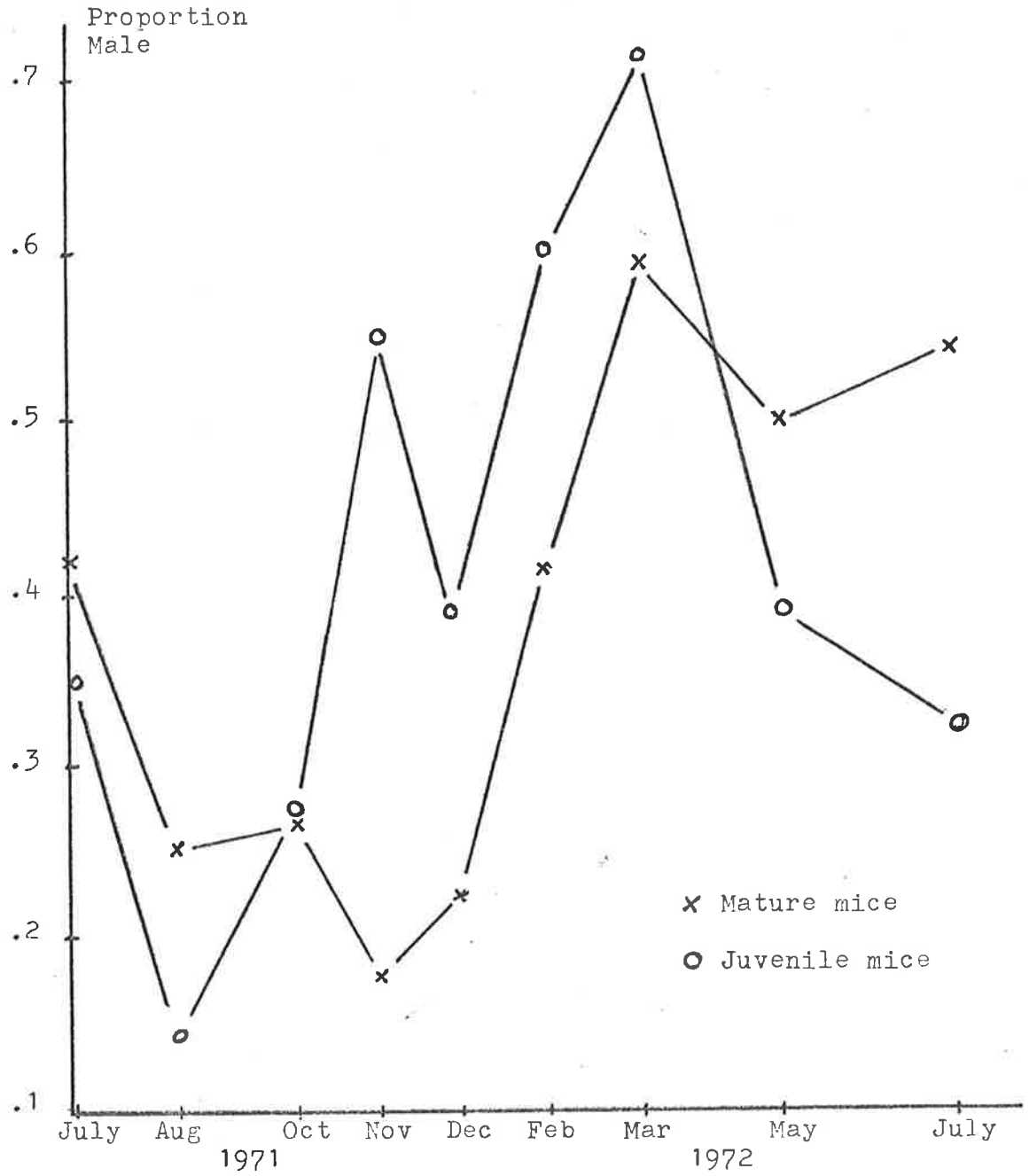
Old Field (OF) and Creek Study Area (CSA) Because the number of traps set and the grid pattern varied during the course of this study it is not reasonable to compare the numbers of mice trapped or estimated on these sites at different times. Instead, the trap rate (mice caught per trap per night) is used as an index of population density (e.g. Brown 1953). The results are given in Figure 10.7 with bars to indicate the range of $\pm$ 2 S.E. The OF colony was small in autumn 1971 when it probably had still not recovered from the crash after the enormous plague of 1970. It gradually increased in the next spring and early summer and then in late summer there was a sharp rise and fall. The intensive trapping study in late March 1972 estimated the population density as 200 per hectare. The CSA colony had a low density through the summer with an increase in the autumn and winter similar to the population fluctuations observed previously. (Newsome 1969b, 1970.)

10.2.4 Reproductive State

Male mice with scrotal testes are usually fertile whilst those with abdominal testes are not (De Long 1967, Newsome 1969a), although in some males the testes readily move between the two locations (Strecker and Emlen, 1953; P.R. Pennycuik, personal communication) and so this is not always a reliable index of fertility. The frequencies of mice in TP and CRK with testes scrotal or abdominal for size classes pooled in pairs are given in Figure 10.8a. There is a linear relationship between the frequency of scrotal testes and HB length in the interval 74mm to 88mm suggesting that the average male developed scrotal testes when about 83mm long or 12 weeks old (Figure 10.5).

Female mice with closed vaginae are usually immature or anoestrous

FIGURE 10.9
SEX RATIO IN TP AND CRK



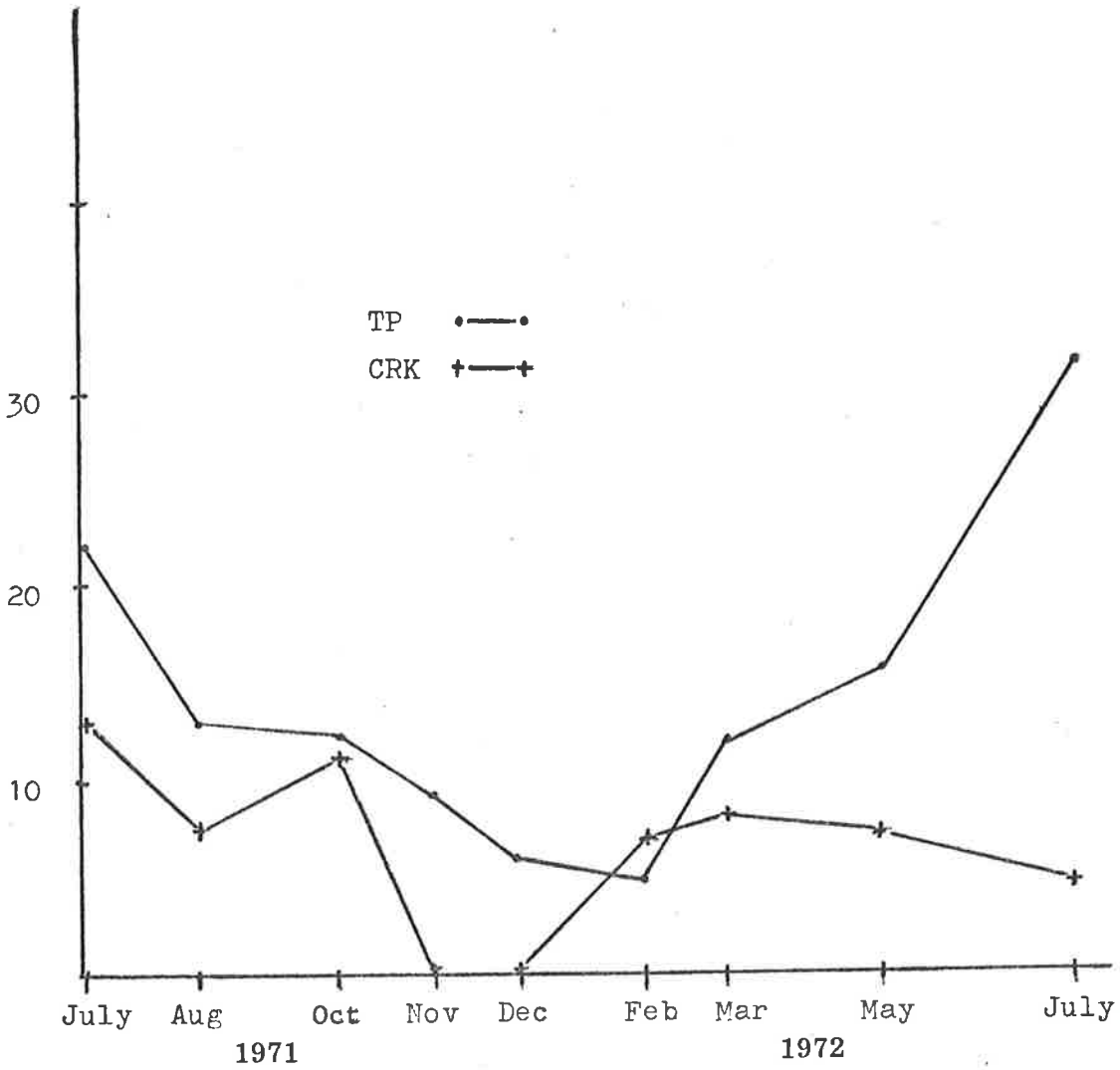
and hence not fecund, whilst those with an open vagina will be estrous (De Long 1967, Newsome 1969a). The frequencies of non-fecund, of fecund but not breeding, and of pregnant or lactating females in TP and CRK for age classes pooled in pairs are presented in Figure 10.8b. The linear decline in the proportion of non-fecund mice suggests that the average female becomes fecund at 85mm or 10 weeks old. These estimates of the average size (or age) at puberty will be taken as the cutoff points for classifying mice as juvenile or mature. Males less than 83mm and females less than 85mm long will be classed as juvenile and all others as mature.

Similar calculations with the less extensive data on mice from the CSA and OF colonies showed that both sexes reached puberty when about 80mm long. The reason for the difference between the feedstation and control colonies is open to speculation. Possibly the higher effective density of mice in the feedstations inhibits sexual maturation, or perhaps the control mice grow more slowly (less well fed) and mature at a smaller size, although the same age as TP and CRK mice.

10.2.5 Sex ratio

Anderson (1970) has stated that in stable mouse colonies there will be a surplus of females among the mature mice. In the feedstation populations the sex ratio has been analysed separately for juvenile and mature mice. Since there was no difference in sex ratio between the TP and CRK colonies at any census, these data were pooled. There was significant heterogeneity between the sex ratios of adult and juvenile mice on only one occasion, November 1971. But when the sex ratio in both age classes is plotted against the time of census (Figure 10.9), there is clearly a cyclical change in sex ratio. The sex ratio changes consistently in juvenile mice before it changes in adults, suggesting that

FIGURE 10.10
Effective Number of mature mice in TP and CRK



regulation of sex ratio occurs during or before the juvenile stage (pre-weaning or foetal).

In the progeny of wild mice bred in the laboratory, there was a slight excess of females at weaning (93♀:80♂) but no difference was found between winter and summer.

10.2.6 Effective number of mature mice

When the sex ratio deviates from 1:1, then the variance or inbreeding effective population size will be less than the actual population size of mature mice. To provide an estimate of the inbreeding effective number (N_e) of mature mice in TP and CRK at each census, taking into account the variation in sex ratio, I have used the formula (Wright, 1931)

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

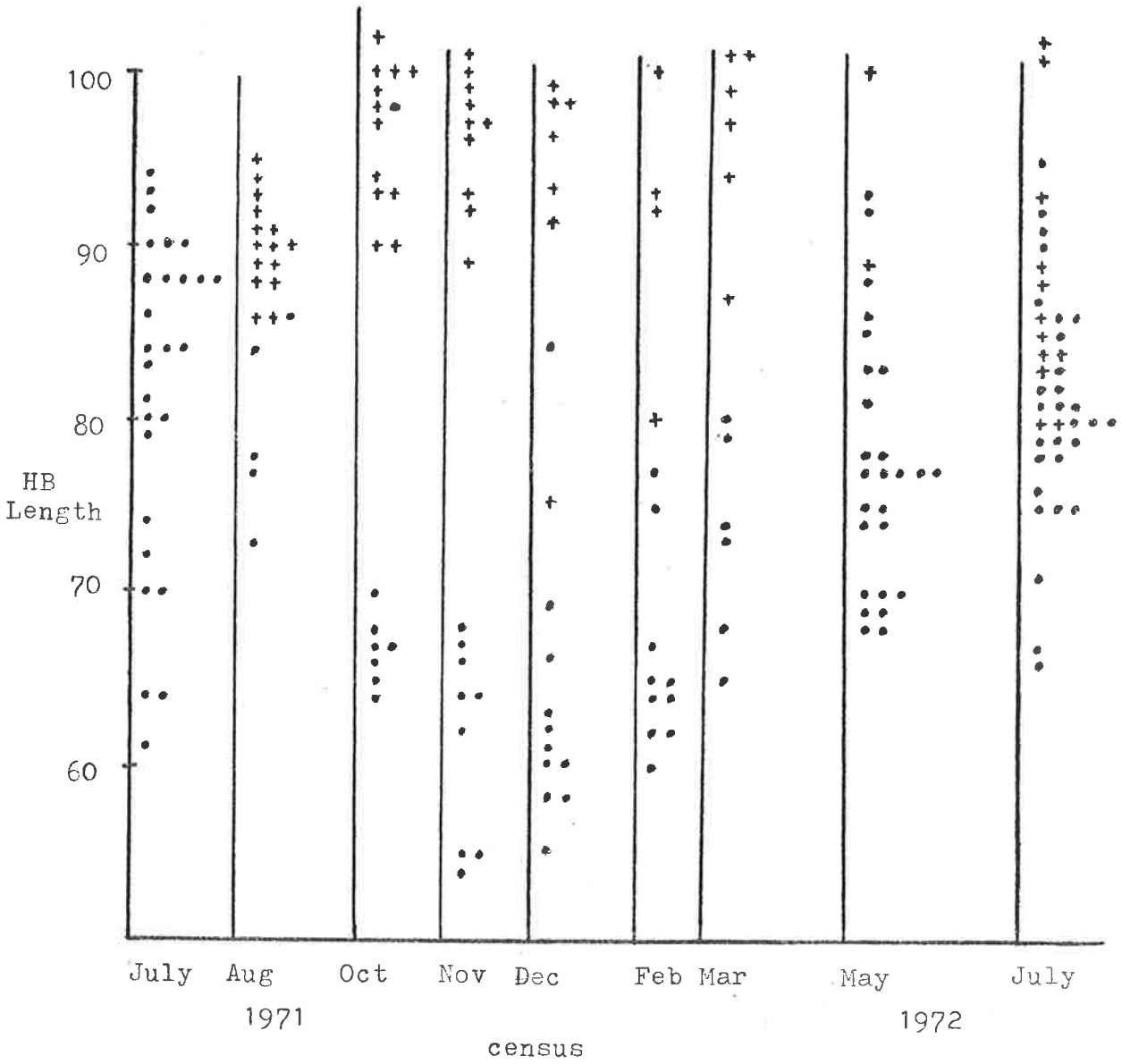
where N_m and N_f are the numbers of the mature males and females respectively.

The results are shown in Figure 10.10. In TP, N_e is small in summer (17% of trappable population) and large in winter (52% to 53% of trappable population in July, 1971 and 1972). The harmonic mean effective population size (the appropriate average when inbreeding effective sizes fluctuate) for the year is 10.4. The CRK colony was never very large and was extinct for over one month. Ignoring the two censuses when CRK appeared to lack a pair of mature adults, the harmonic mean effective number was 7.7. Since the Lincoln index data suggest that only 83% of mice are caught each census, these population sizes can be increased by 25% to compensate for this. Thus the average effective size of TP is 13 and of CRK is 9.6 per census, which is about two mice per feedstation.

10.2.7 Size Structure of TP colony.

If the age-structure of this colony had remained approximately

FIGURE 10.11
Size Distribution of Female mice in TP



• First time trapped

+ Trapped previously

FIGURE 10.12

Size Distribution of male mice in TP

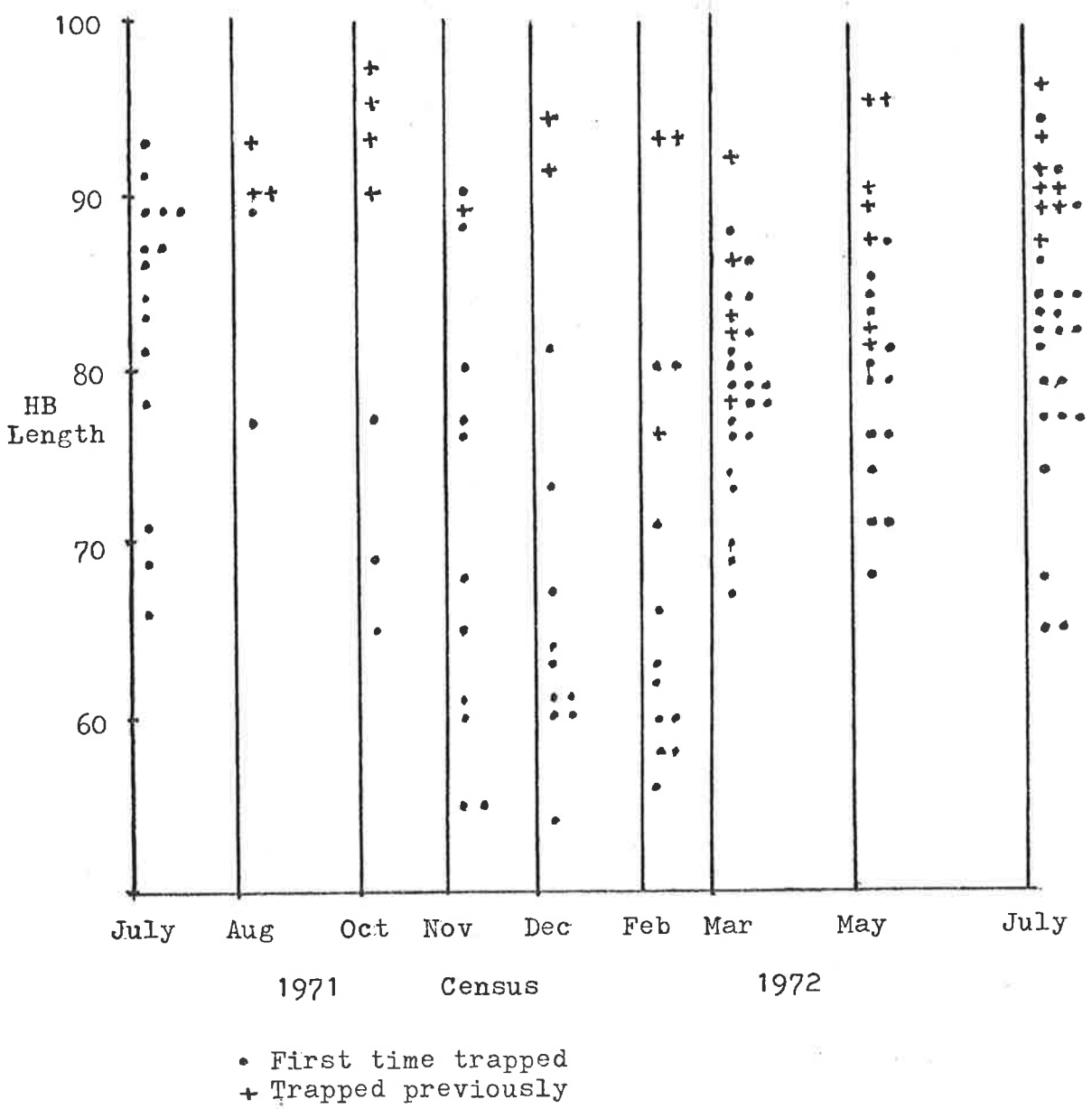


TABLE 10.1 Mean Age of Mature Mice in TP colony

Census	♀ age in weeks	♂ age in weeks
30.8.71	15.06 (16)*	17.75 (4)*
18.10.71	19.15 (13)	27.00 (4)
22.11.71	21.30 (10)	22.30 (3)
20.12.71	22.30 (6)	43.00 (2)
2.2.72	30.00 (3)	28.00 (2)
4.3.72	30.67 (6)	16.33 (6)
29.4.72	18.14 (7)	20.78 (9)
7.7.72	23.93 (15)	20.35 (17)
Mean	21.01 (76)	21.30 (47)

* Sample size

constant through the year, it would have been valid to derive life tables from which the reproductive value (Fisher, 1958), generation time and effective population size could be calculated. However, the size distribution at each census in TP (Figures 10.11a and 10.11b) is clearly dependent upon seasonal factors, being unimodal in winter and bimodal in summer.

10.2.8 Generation time of TP colony

The generation time (or length) of a population with overlapping generations is given by a weighted mean age of reproducing individuals: each age class is weighted by its proportional contribution to the amount of reproduction in the population. If it can be assumed that all mature individuals contribute equally (i.e. there are no significant effects of age on reproduction once maturity is achieved), then the average age of mature individuals is the same as the generation length. The relationship between age and reproductive success is unknown for the mice in this study, so the mean age of mature mice will be taken as an approximation to the generation length. For mice in the TP colony, the age at first capture can be estimated from the growth curve in Figure 10.5, providing they are younger than 16 weeks, otherwise they can be compared with mice of a similar size and same sex that have been 'aged' at an earlier census. The July 1971 census contained many mice that could not be aged properly so this census was ignored. The mean estimated ages for males and females are given in Table 10.1. Although the mean age in both sexes is 21 weeks, there is some seasonal variation. Since mice are conceived three weeks before birth, this must be added on to the mean ages to given an estimate of 24 weeks for the generation time i.e. there are two generations per year.

10.2.9 Recapture rates for different age classes in TP.

The mice were grouped into three age classes (3-9 weeks, 9-15 weeks,

TABLE 10.2 Recapture Rates for Three Age-classes in the TP Population

Sex	Age Classes		
	3-9 weeks	9-15 weeks	⩾ 15 weeks
♀	.37 ± .06	.76 ± .09	.56 ± .07
♂	.19 ± .05	.25 ± .11	.64 ± .09

TABLE 10.3 Recapture in same census in TP.

Feedstation of last capture	Feedstation of recapture (each recapture scored separately)					
	1	2	3	4	5	6
FEMALES						
1	30	3				
2	2	31				
3		2	14			
4				55	1	
5				1	33	
6						16
MALES						
1	15	2				
2	4	14	1			
3	1	1	18			
4				12	1	1
5					8	
6				1		21

over 15 weeks) and the recapture rate (i.e. proportion recaught at a subsequent census) calculated for each age class in each sex (Table 10.2). The two sexes are significantly different only in the 9-15 weeks age class. Each sex is significantly heterogeneous over the three age classes. Males have a lower recapture rate (implying a lower survival rate) than females until after puberty. These recapture rates cannot validly be used as estimates of survival rates because the intervals between censuses differ and survival rates appear to vary with the seasons. As an exercise, the recapture rates were used to estimate average survival rates per six weeks epoch and a life table was calculated on the assumption that the population of trappable mice was 36. However, the results did not resemble TP in generation time,sex ratio or age structure and merely served to confirm that the approach was invalid.

10.2.10 Movement of mice

The movement of mice between captures can be examined at two levels, within each census and between censuses. The data on the site of recapture of a mouse already caught one or more times during a census in TP and CRK shows that 93.7%($\pm$ 1.5%) of female recaptures and 86.7%($\pm$ 2.7%) of male recaptures were made at the same feedstation as that where they were last caught. The difference between the two sexes is statistically significant. ($2.5\% < p < 5\%$.) (The data for TP are given in Table 10.3.)

If the recapture of mice that were always or were predominantly caught under a particular feedstation during a census is studied, it is found that at the next census when those mice are caught, the proportion that are always recaught in the same feedstation as before is 81$\pm$4% for females and 56.5$\pm$7% for males. (Probability for difference <1%.) All 13 of the males and five of the eight females which changed feedstations within a census were mature mice. These data suggest that female mice are more restricted in their movements than male mice. An explanation

TABLE 10.4 Site of Recapture of Female Mice in all Censuses After First Capture.

Site of First Capture	Site of subsequent recapture			
	TPA	TPB	CRK	OF
TPA	14	1	1	
TPB	1	16		
CRK			17	
OF				9
Observed immigration	1	1	1	0
	15	17	18	9

for the greater mobility of males is that their density is often less than one per feedstation and as each feedstation should receive the attention of at least one male, two or more feedstations may be included in the home range of a mature male. Home ranges are not necessarily exclusive as it was not uncommon to trap two mature males under the same feedstation on the same night.

Migration between colonies can be estimated by comparing the number of mice recaught at a different colony from the one in which they were initially trapped, with the number of mice recaught in the same colony. In this section, TP will be split into two colonies, TPA and TPB, both with three feedstations. None of the 37 males which were recaught at least once changed colonies whilst three of 50 females did so, one going from TPA to TPB, one going the other way and one going from TPA to CRK, Table 10.4. No mouse was observed to go to or from OF or CSA, but as sample sizes were small in these two colonies, this is not surprising. The known female immigration rate into TPA and TPB is 2/32=.0625 and into CRK is 1/18=.055. These are minimum estimates of the actual immigration rates because it is unlikely that all mice were marked before they were old enough to disperse and not all mouse colonies in the vicinity (especially along Salt Creek) were trapped. The immigration rate into CRK after the early summer crash would have been large as not enough mice remained to re-establish it on their own. The females that migrated were young mice, two were only recently weaned (HB=64mm) and one was near puberty (84mm) when first caught and they had changed colonies by the next census.

Predominantly young mice were caught at traps set some distance from the feedstations in TP and CRK. In TP, 10 out of 10 females and 14 out of 19 males caught away from the feedstations were juvenile whilst in CRK, 10 of 13 females and six of 10 males were juvenile. Some of the

TABLE 10.5 Maximum Distance between Capture Sites for mice in the Oldfield in March, 1972.

Sex	Distance in metres						
	0	8.22	9.14	12.30	16.43	18.3-20.1	⩾20.1
♀	14	2	3	3	0	0	0
♂	10	1	3	1	1	5	2

mature mice in CRK were from the October 1971 census, which was the only occasion when mice were observed breeding outside of the feedstations in TP or CRK. The immaturity of these mice suggests that juvenile mice are driven out of the established colonies living under feedstations and they either die, help to found new colonies or manage to re-establish themselves in one of the old colonies.

10.2.11 OF Intensive study.

83 females and 87 males were caught in this study. Known mortality was 9.4% and the Lincoln index calculated from the results of the last night gave an estimate of 227 as the population size i.e. a density of 200/ hectare. For the mice recaught at least once, the maximum distance between capture sites is given in Table 10.5. There is little long distance movement but recaptured males sometimes move further than females.

10.3 GENETICAL RESULTS

10.3.1 Introduction

Each mouse caught was scored for the *Agouti* locus and if a blood sample could be taken it was also scored for *Es-1*, *Es-3*, *Hbb* and *Erp-1*. Sample sizes for the electrophoretically detectable loci are usually smaller than for the *Agouti* locus because some mice died before being bled and sometimes a blood sample was too small for it to be scored for all loci. Statistical analyses (such as those involving χ^2) which are based on the assumption that samples have been drawn from large populations are not appropriate to these data (Smith, 1969; Neel and Ward, 1972). From each colony, about 80% of the trappable mice have been scored and hence the sampling variances are much less than they would be if the samples came from very large populations. Thus the genetic differences between colonies will be studied by F-statistics again, instead of χ^2.

Gene frequencies will be calculated from all the mice caught in a census of a colony without making a correction for known deaths. These frequency estimates are thus of the genes present in a colony at the start of each census. Sometimes, mice which were not caught on a particular census, were caught on previous and subsequent censuses. This suggests that they were in fact present in the colony, but were not caught. Whilst it is common practice to include such mice in the census in which they were missed (e.g. Tamarin and Krebs, 1971; Gaines and Krebs, 1971), I did not do this because this procedure introduces an additional source of error into gene frequency estimates. The increase in sampling variance is difficult to predict, and this is a drawback that outweighs the small increase in sample size obtained by including such absent mice.

Very often, juvenile mice in a colony could be identified as members of the same litter on the criteria of similar size, similar

TABLE 10.6 Phenotype numbers at the *Agouti* locus

Sample	A^{w}.	$A^{+}A^{+}$	Total	Estimated frequency of A^{+} F=0	F=-.12
TPA1	21	0	21	0.000	.107
TPB1	8	13	21	.787	.799
CRK1	25	2	27	.272	.316
TOT1	54	15	69	.466	
TPA2	15	0	15	0.000	.107
TPB2	4	7	11	.798	.809
CRK2	9	1	10	.316	.357
TOT2	28	8	36	.471	
TPA3	14	0	14	0.000	.107
TPB3	6	7	13	.734	.749
CRK3	18	6	24	.500	.529
TOT3	38	13	51	.505	
TPA4	12	0	12	0.000	.107
TPB4	7	11	18	.782	.794
CRK4	6	0	6	0.000	.107
TOT4	25	11	36	.553	
TPA5	15	0	15	0.000	.107
TPB5	7	7	14	.707	.724
CRK5	7	0	7	0.000	.107
TOT5	29	7	36	.441	
TOT6	36	6	42	.378	
TOT7	52	7	59	.344	
TPA8	19	1	20	.224	.272
TPB8	17	8	25	.566	.591
CRK8	30	2	32	.250	.296
TOT8	66	11	77	.378	
TPA9	32	0	32	0.000	.107
TPB9	21	9	30	.548	.574
CRK9	19	0	19	0.000	.107
TOT9	72	9	81	.333	

TABLE 10.7 Genotype numbers at the *Es-1* locus

Sample	*bb*	*ab*	*aa*	Total (N)	Frequency of *Es-1*^{b}
TPA1	19	0	0	19	1.000
TPB1	15	3	0	18	.917 ± .044*
CRK1	20	4	0	24	.917 ± .038
TOT1	54	7	0	61	.943 ± .021**
TPA2	13	0	0	13	1.000
TPB2	10	1	0	11	.955 ± .043
CRK2	8	2	0	10	.900 ± .063
TOT2	31	3	0	34	.956 ± .025
TPA3	14	0	0	14	1.000
TPB3	11	1	0	12	.958 ± .040
CRK3	21	2	0	23	.957 ± .029
TOT3	46	3	0	49	.969 ± .017
TPA4	11	0	0	11	1.000
TPB4	15	1	0	16	.969 ± .030
CRK4	6	0	0	6	1.000
TOT4	32	1	0	33	.985 ± .015
TPA5	12	3	0	15	.900 ± .052
TPB5	12	1	0	13	.962 ± .037
CRK5	3	4	0	7	.714 ± .094
TOT5	27	8	0	35	.886 ± .038
TOT6	27	11	0	38	.855 ± .040
TOT7	43	15	0	68	.871 ± .031
TPA8	16	4	0	20	.900 ± .045
TPB8	19	6	0	25	.880 ± .043
CRK8	23	8	1	32	.844 ± .047
TOT8	58	18	1	77	.870 ± .027
TPA9	26	4	0	30	.933 ± .031
TPB9	22	7	0	29	.879 ± .040
CRK9	15	3	0	18	.917 ± .044
TOT9	63	14	0	77	.909 ± .023

* S.E. = $q(1-q)(1+F)/2N$

** S.E. = $q(1-q)/2N$

TABLE 10.8 Genotype numbers at the *Es-3* locus

Sample	*cc*	*bc*	*bb*	Total (N)	Frequency of $Es\text{-}3^c$
TPA1	19	1	0	20	.975 ± .024*
TPB1	5	13	0	18	.639 ± .053
CRK1	15	9	0	24	.813 ± .049
TOT1	39	23	0	62	.815 ± .035**
TPA2	13	0	0	13	1.000
TPB2	4	5	2	11	.591 ± .108
CRK2	7	3	0	10	.850 ± .072
TOT2	24	8	2	34	.824 ± .046
TPA3	14	0	0	14	1.000
TPB3	5	6	1	12	.667 ± .090
CRK3	19	3	1	23	.891 ± .053
TOT3	38	9	2	49	.867 ± .034
TPA4	9	2	0	11	.909 ± .058
TPB4	8	6	2	16	.688 ± .087
CRK4	4	2	0	6	.833 ± .096
TOT4	21	10	2	33	.788 ± .050
TPA5	7	8	0	15	.733 ± .064
TPB5	6	8	0	14	.714 ± .066
CRK5	1	3	3	7	.357 ± .132
TOT5	14	19	3	36	.653 ± .056
TOT6	16	20	2	38	.684 ± .053
TOT7	23	30	5	58	.655 ± .044
TPA8	8	12	0	20	.700 ± .055
TPB8	12	12	1	25	.720 ± .057
CRK8	15	14	3	32	.688 ± .057
TOT8	35	38	4	77	.695 ± .037
TPA9	10	18	2	30	.633 ± .052
TPB9	14	10	5	29	.655 ± .069
CRK9	5	11	2	18	.583 ± .071
TOT9	29	39	9	77	.630 ± .037

* S.E. = $q(1-q)(1+F)/2N$

** S.E. = $q(1-q)/2N$

TABLE 10.9 Genotype numbers at the *Hbb* locus

Sample	*ss*	*sd*	*dd*	Total (N)	Frequency of Hbb^s
TPA1	17	3	0	20	.925 ± .040*
TPB1	17	1	0	18	.972 ± .027
CRK1	11	8	5	24	.625 ± .079
TOT1	45	12	5	62	.823 ± .034**
TPA2	11	2	0	13	.923 ± .050
TPB2	10	1	0	11	.955 ± .043
CRK2	5	2	3	10	.600 ± .138
TOT2	26	5	3	34	.838 ± .045
TPA3	14	0	0	14	1.000
TPB3	11	1	0	12	.958 ± .040
CRK3	10	3	6	19	.605 ± .102
TOT3	35	4	6	45	.822 ± .040
TPA4	11	0	0	11	1.000
TPB4	16	0	0	16	1.000
CRK4	2	3	1	6	.583 ± .140
TOT4	29	3	1	33	.924 ± .032
TPA5	6	9	0	15	.700 ± .063
TPB5	13	1	0	14	.964 ± .034
CRK5	7	0	0	7	1.000
TOT5	26	10	0	36	.861 ± .041
TOT6	20	17	1	38	.750 ± .050
TOT7	37	20	0	57	.825 ± .036
TPA8	16	4	0	20	.900 ± .045
TPB8	14	10	1	25	.760 ± .057
CRK8	22	9	1	32	.828 ± .047
TOT8	52	23	2	77	.825 ± .031
TPA9	22	8	0	30	.867 ± .040
TPB9	20	9	0	29	.845 ± .043
CRK9	6	11	0	17	.677 ± .058
TOT9	48	28	0	76	.816 ± .031

* S.E. = $q(1-q)(1+F)/2N$

** S.E. = $q(1-q)/2N$

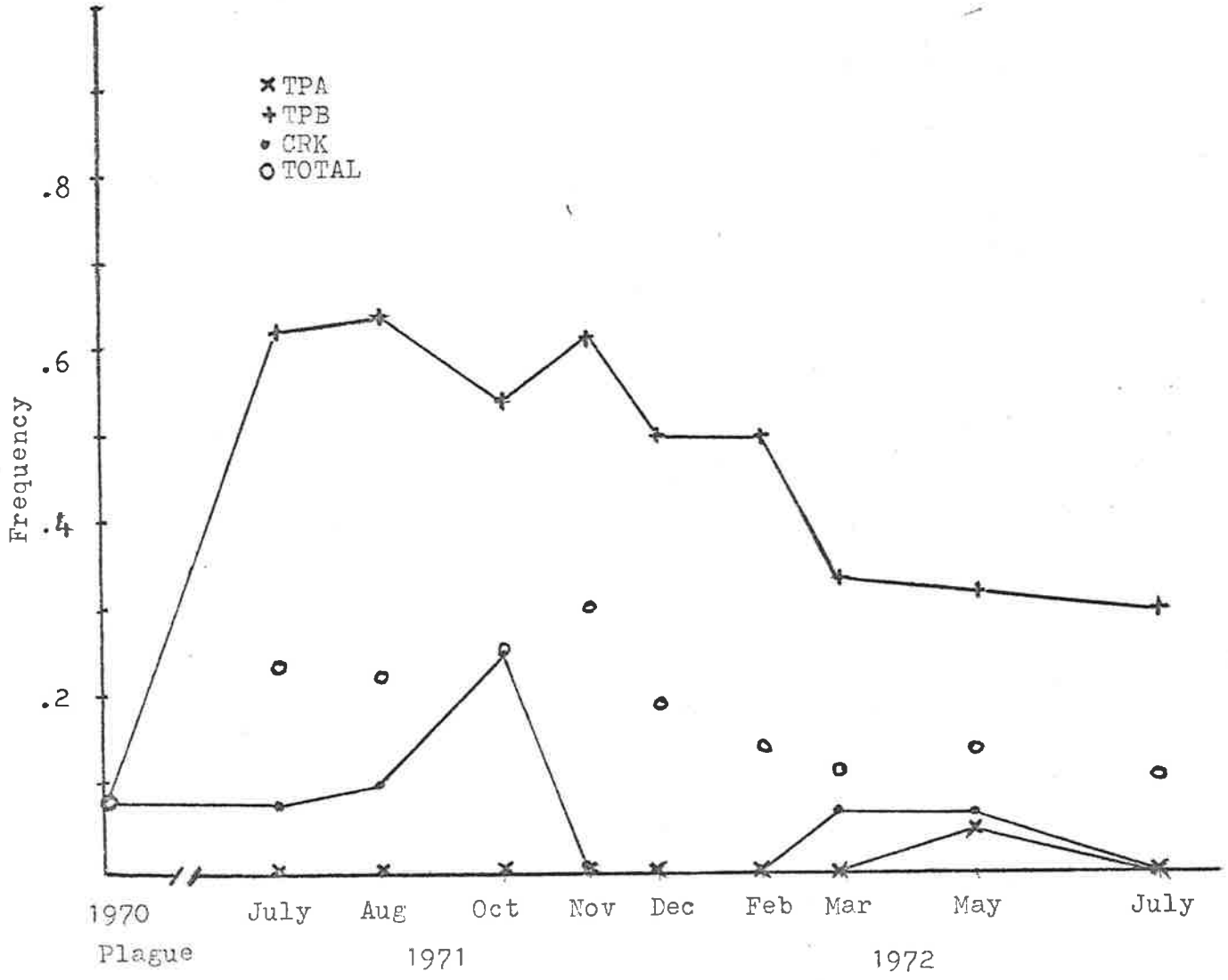
TABLE 10.10 Genotype numbers at the *Erp-1* locus

Sample	*aa*	*ab*	*bb*	Total (N)	Frequency of *Erp-1*[a]
TPA1	19	0	0	19	1.000
TPB1	18	0	0	18	1.000
CRK1	23	1	0	24	.979 ± .020*
TOT1	60	1	0	61	.992 ± .008**
TPA2	13	0	0	13	1.000
TPB2	11	0	0	11	1.000
CRK2	10	0	0	10	1.000
TOT2	34	0	0	34	1.000
TPA3	14	0	0	14	1.000
TPB3	12	0	0	12	1.000
CRK3	18	0	0	18	1.000
TOT3	44	0	0	44	1.000
TPA4	11	0	0	11	1.000
TPB4	16	0	0	16	1.000
CRK4	6	0	0	6	1.000
TOT4	33	0	0	33	1.000
TPA5	15	0	0	15	1.000
TPB5	13	0	0	13	1.000
CRK5	7	0	0	7	1.000
TOT5	35	0	0	35	1.000
TOT6	39	0	0	39	1.000
TOT7	56	1	0	57	.991 ± .009
TPA8	15	5	0	20	.875 ± .048
TPB8	24	1	0	25	.980 ± .020
CRK8	29	3	0	32	.953 ± .026
TOT8	68	9	0	77	.942 ± .019
TPA9	25	5	0	30	.917 ± .034
TPB9	28	1	0	29	.983 ± .017
CRK9	17	0	0	17	1.000
TOT9	70	6	0	76	.961 ± .016

* S.E. = $q(1-q)(1+F)/2N$

** S.E. = $q(1-q)/2N$

FIGURE 10.13
FREQUENCY OF GREY-BELLIED PHENOTYPE



genotypes and capture under the same feedstation.

10.3.2 Phenotype and gene frequencies at each census

In this section the genetic history of each colony will be presented in terms of the frequency of codominant genes and the frequency of recessive phenotypes. Full data on genotypes and phenotypes present at each census in TPA, TPB and CRK are given in Tables 10.6 to 10.10 except for the 6th and 7th censuses (February and March, 1972), for which the data can be obtained from Tables 9.2 to 9.6. Because the OF data are not used in this chapter (too few mice were obtained on most censuses) but are used in Chapter 9, the totals for censuses 6 and 7 ("TOT6" and "TOT7") including only the TPA, TPB and CRK data are also given in Tables 10.6 to 10.10. The standard errors for gene frequencies of codominant genes were calculated using Equation (3.16B) for single colonies and Equation (2.1) for census totals because these equations were thought appropriate.

10.3.2.1 *A* locus

The change in the frequency of the grey-belly (A^+A^+) phenotype in each colony and in the total of the three colonies for the period 1970-2 is presented in Figure 10.13 (see Table 10.6). Three features of this figure should be noted. Firstly the frequency of the grey-belly phenotype is low on average and in individual colonies the phenotype is often missing (e.g. TPA). In CRK it was lost during the population crash (censuses 4, 5 and 6), reappeared after recolonisation and was absent again in the last census. Secondly, the adjacent colonies, TPA and TPB, have very different frequencies of this phenotype. Grey-bellied mice were always common in TPB, but only one was ever caught in TPA. Thirdly, in July 1971, the mean frequency of grey-bellied mice was about three times larger than in 1970 (Table 9.6), but as the study progressed until July 1972, the frequency declined to the earlier value, mainly because of the steady change in TPB.

FIGURE 10.14
FREQUENCY OF Es-1^b AND Erp-1^b GENES

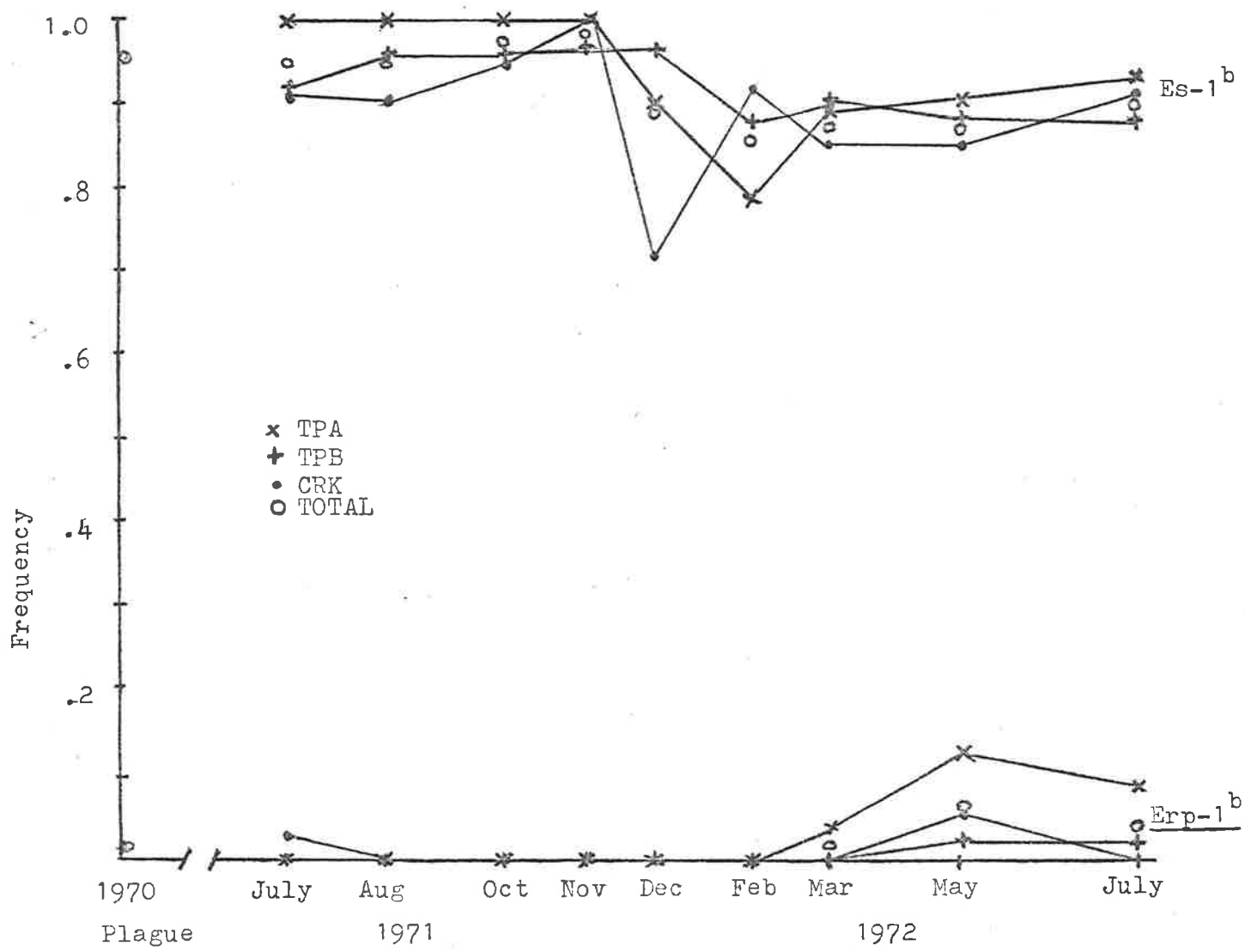
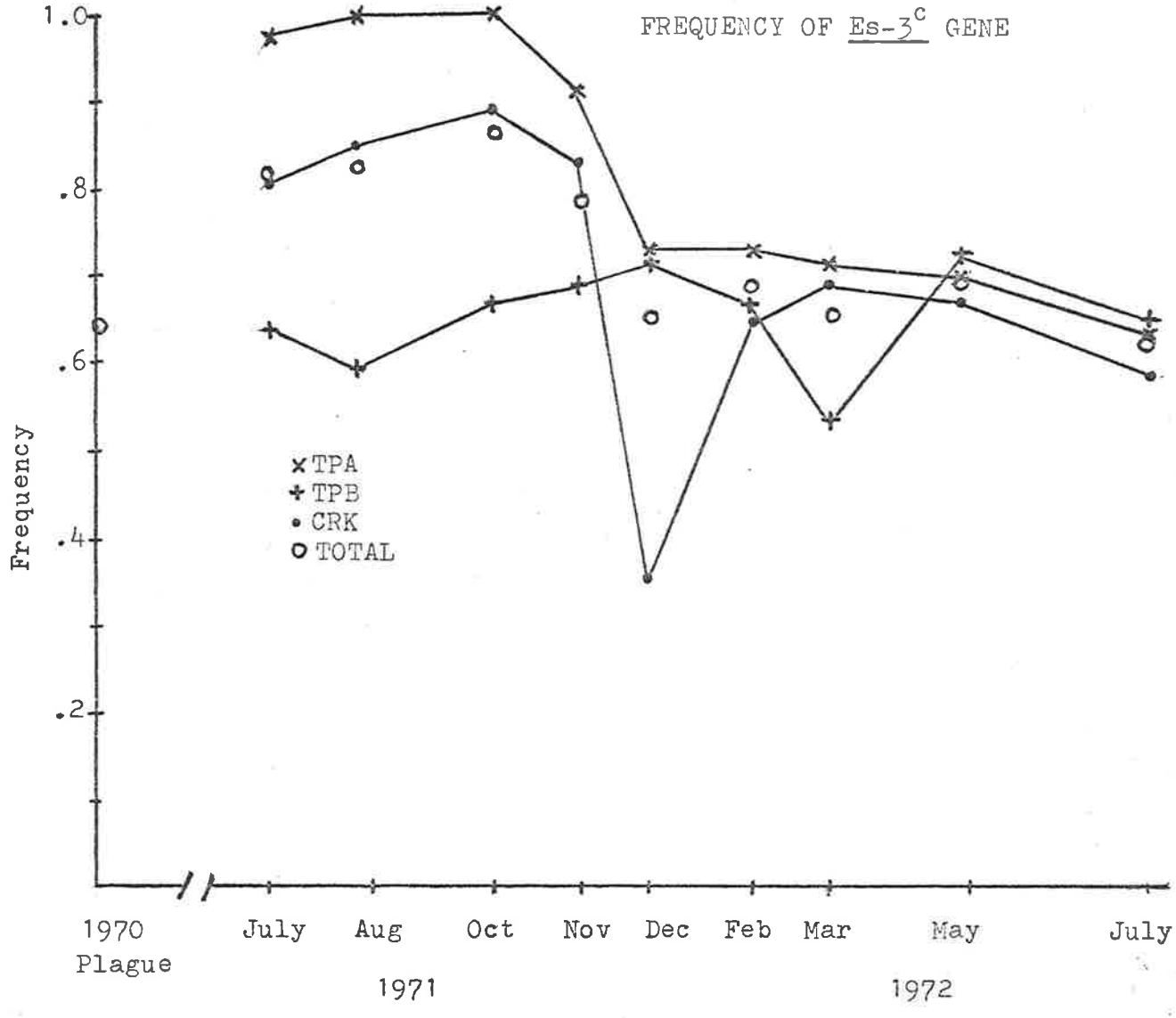


FIGURE 10.15
FREQUENCY OF Es-3^c GENE



10.3.2.2 *Es-1* locus

The frequency of the $Es\text{-}1^b$ allele in all three colonies and in the total is depicted in Figure 10.14 (see also Table 10.7). From the figure it can be seen this allele was fixed (TPA) or almost fixed (TPB and CRK) in the colonies until censuses 5 and 6. The fluctuations in frequency in CRK between censuses 4, 5, 6 and 7 can be ascribed to the effects of small population size during the population crash in this colony.

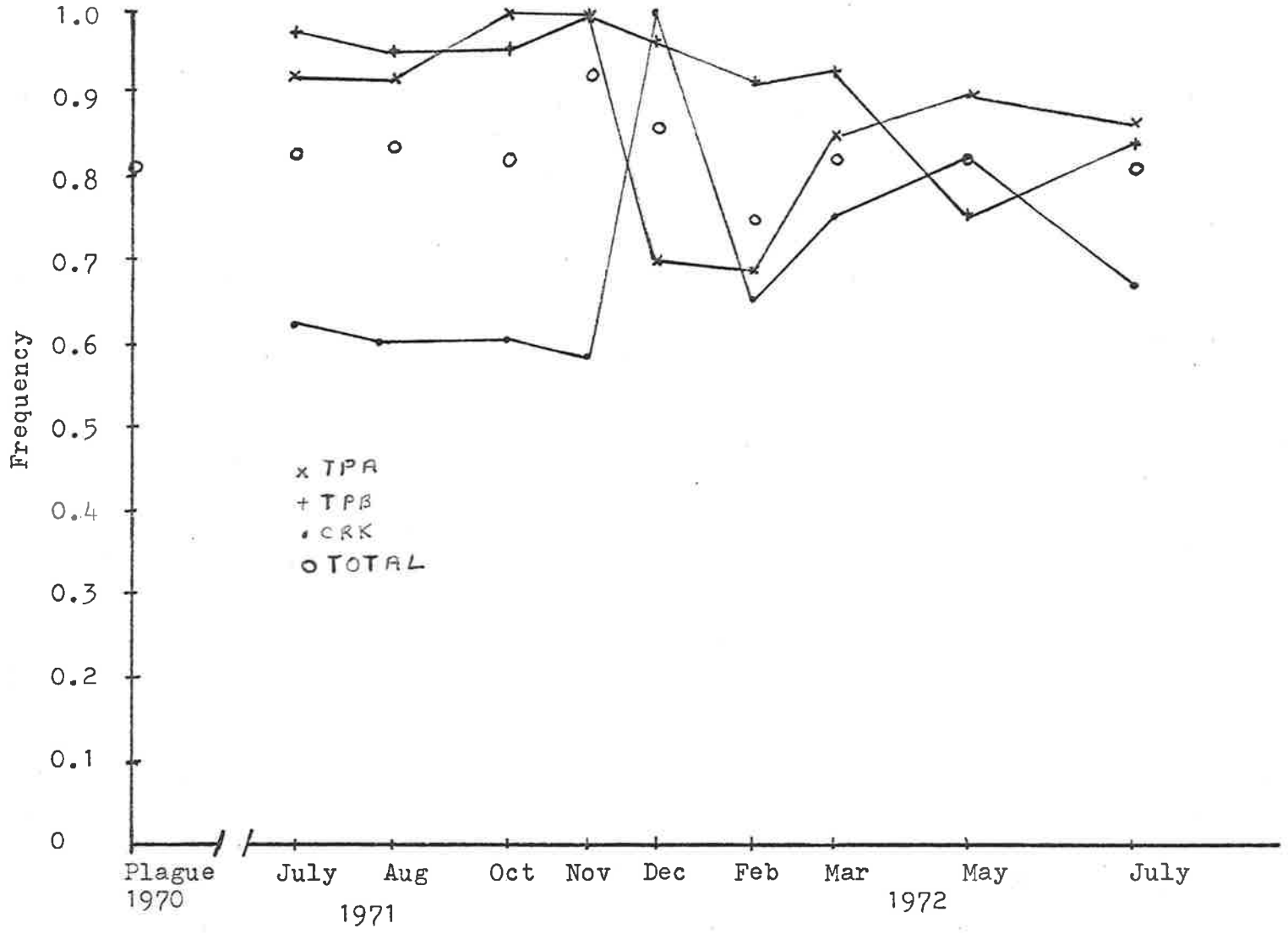
The decrease in frequency of $Es\text{-}1^b$ in TPA in censuses 5 and 6 appears to be due to a heterozygous male, because three of ten juveniles in census 5 and six of eight juveniles in census 6 in feedstations 1 and 2 in TPA were heterozygous $Es\text{-}1^a\ Es\text{-}1^b$. Since all adult mice caught in TPA in censuses 1 to 6 were $Es\text{-}1^b\ Es\text{-}1^b$, the most likely explanation for the sudden appearance of so many $Es\text{-}1^a\ Es\text{-}1^b$ juveniles in two feedstations is that an $Es\text{-}1^a\ Es\text{-}1^b$ male immigrated and became dominant under feedstations 1 and 2. Further genetic evidence about this mouse will be given below. Heterozygous ($Es\text{-}1^a\ Es\text{-}1^b$) juveniles were not caught under the third feedstation in censuses 5, 6 and 7, suggesting that a different male was dominant in this feedstation.

10.3.2.3 *Es-3* locus

From Figure 10.15 (see also Table 10.8) it can be seen that in July 1971, the frequencies of the $Es\text{-}3^c$ gene were very different in the three colonies, but by the end of the study period (July 1972), frequencies were very similar. Also, the mean gene frequency in July 1972, is similar to that of the 1970 plague. The gene frequency in TPB oscillated about the 1970 value while the frequency in TPA converged to it midway through the study period. CRK underwent drastic fluctuations in gene frequency between censuses 4, 5 and 6, again because of the small numbers present in the colony during that period.

FIGURE 10.16

FREQUENCY OF Hbb^S GENE



Genetical evidence from the *Es-3* locus suggests that the TPA colony received at least two immigrant males. In census 1, the only heterozygous ($Es\text{-}3^b$ $Es\text{-}3^c$) mouse died and no more heterozygotes were observed until census 4 when two male heterozygotes (one mature and one juvenile) were caught in feedstation 3. It is likely that these two were immigrants. One of these, or another heterozygous male, presumably became established under feedstation 3 since the only adult female caught under this feedstation from census 3 to census 7 was $Es\text{-}3^c$ $Es\text{-}3^c$ and yet 9 of 15 juveniles in censuses 5, 6, 7 and 8 were $Es\text{-}3^b$ $Es\text{-}3^c$ heterozygotes. This observation is evidence for there being at least one reproductively successful immigrant in feedstation 3.

Under feedstations 1 and 2 in TPA, many of the juvenile mice in censuses 5, 6 and 7 were also heterozygous ($Es\text{-}3^b$ $Es\text{-}3^c$). These juveniles were also often $Es\text{-}1^a$ $Es\text{-}1^b$, showing that they were the progeny of the $Es\text{-}1^a$ $Es\text{-}1^b$ male immigrant discussed in 10.3.2.2. All five mature females caught in feedstations 1 and 2 in censuses 4, 5, 6 and 7 were $Es\text{-}3^c$ $Es\text{-}3^c$, so the $Es\text{-}3^b$ gene had to come from an immigrant male. Thus there is good evidence that there were at least two male immigrants: one went to feedstation 3 and the other to feedstations 1 and 2.

10.3.2.4 *Hbb* locus

The frequency of Hbb^s is shown in Figure 10.16 (see Table 10.9), where it can be seen that the average frequency remained remarkably similar throughout this study to the frequency observed in 1970. CRK again shows dramatic gene frequency changes during its population crash. The apparent fixation of Hbb^s in TPB in census 4 is probably a sampling error: a mature Hbb^s Hbb^d female caught in feedstation 6 in censuses 3, 5, 7, 8 and 9 was probably also present in census 4 but I failed to catch her. This heterozygous female (Hbb^s Hbb^d) was possibly responsible for part or all of the decline in the frequency of Hbb^s in the TPB colony because there were 10 Hbb^s Hbb^d mice out of 22 juveniles

caught in feedstation 6 in censuses 6, 7, 8 and 9. No mouse carrying the Hbb^d allele was observed in feedstations 4 and 5 until census 8. These Hbb^s Hbb^d mice in feedstations 4 and 5 could have been the descendants of the above mentioned female and need not be immigrants from outside TPB.

In TPA, there was a dramatic change in the frequency of Hbb^s between censuses 4 and 5 because all of the juveniles caught in feedstations 1 and 2 in censuses 5, 6 and 7 were Hbb^d Hbb^s. Since all these juveniles were also segregating at the *Es-1* and *Es-3* loci, the same male was also the father of all the progeny discussed in the sections on *Es-1* and *Es-3*. I have now assembled evidence from three loci to show that a single male was dominant under feedstations 1 and 2 in TPA and that his genotype was $Es\text{-}1^a$ $Es\text{-}1^b$, $Es\text{-}3^b$ $Es\text{-}3^c$, Hbb^d Hbb^d. Because of his distinctive genotype, all of his presumed progeny (at least 21 were caught) can be identified. Also, because he carried three genes that were rare or absent from the colony before he appeared, he is certainly an immigrant. In addition, it is very unlikely that there were two immigrant males of similar genotype because all the juveniles were Hbb^s Hbb^d, and all litters were segregating at the *Es-1* and *Es-3* loci.

By estimating the age of his progeny, it is possible to extrapolate back to find their approximate conception dates. From these dates I deduce that he was active during censuses 3, 4 and 5, and perhaps even census 6 (judging by a small litter of three in feedstation 1 in census 8). But no mouse of his genotype was ever caught, demonstrating that the trapping program failed to catch all of the mature mice in the TPA colony.

Only two of the 21 definite progeny of this presumed immigrant male caught in censuses 5, 6 and 7 were ever recaught at a later census. It should be noted that they were raised during a period when most

juvenile mice were being eliminated from the colony before maturity. If these progeny had been born later (in the autumn) they would have been part of the cohort that formed the overwintering population. In this case, the effects of that particular male on gene frequencies in the TPA colony would have been even more pronounced.

10.3.2.5 *Erp-1* locus

For much of the study period most colonies showed no variation at the *Erp-1* locus (Figure 10.14, Table 10.10). As $Erp\text{-}1^b$ is a rare gene, it is expected to be absent from most small colonies of mice, but in some colonies it may, by chance, reach moderate frequencies. This drift to moderate frequencies is shown in TPA in censuses 8 and 9 when the frequency of $Erp\text{-}1^b$ reached 10 per cent.

One juvenile heterozygote was caught in feedstation 2 (TPA) in census 7 and the same one plus four more were caught in census 8. From the sizes of these heterozygotes it is not clear if there were one or two litters, so the heterozygous parent (probably an immigrant) could have been male or female. If there were two litters of similar ages, the heterozygous parent would have to be a male because it is very unlikely that two heterozygous females could have been breeding at the same time when none had been seen before in that colony.

In census 8, a heterozygote also appeared in TPB (feedstation 5) and judging by its phenotype at the *A*, *Es-1* and *Es-3* loci and its size, it could have come from the litter(s) in feedstation 3 (TPA). Whatever its origin, this mouse was an immigrant into TPB.

10.3.3 Genotype frequencies

The estimated sample fixation indices (F_{IS}) for the *Erp-1*, *Es-1*, *Es-3* and *Hbb* loci in all censuses are presented in Table 10.11 (see also Table 9.7). The fixation indices for the genotype data pooled over all three colonies (denoted "TOT1", etc.) are also given: these represent estimates of F_{IT} in the population made up by the three colonies. The

TABLE 10.11 Sample Fixation indices (F_{IS})

Sample	*Hbb*	*Es-3*	*Es-1*	*Erp-1*
TPA1	-.081 ± .047	-.026 ± .026	-	-
TPB1	-.029 ± .029	-.565 ± .129	-.091 ± .052	-
CRK1	.289 ± .200	-.231 ± .075	-.091 ± .045	-.021 ± .021
TOT1	.337 ± .151	-.228 ± .046	-.061 ± .023	-.008 ± .008
TPA2	-.083 ± .059	-	-	-
TPB2	-.048 ± .048	.060 ± .302	-.048 ± .048	-
CRK2	.583 ± .262	-.177 ± .100	-.111 ± .078	-
TOT2	.458 ± .206	.191 ± .200	-.046 ± .027	-
TPA3	-	-	-	-
TPB3	-.044 ± .043	-.125 ± .275	-.044 ± .043	-
CRK3	.669 ± .175	.327 ± .299	-.046 ± .032	-
TOT3	.696 ± .142	.202 ± .183	-.032 ± .018	-
TPA4	-	-.100 ± .070	-	-
TPB4	-	.127 ± .256	-.032 ± .032	-
CRK4	-.029 ± .407	-.200 ± .139	-	-
TOT4	.351 ± .292	.093 ± .187	-.015 ± .015	-
TPA5	-.429 ± .129	-.364 ± .120	-.111 ± .064	-
TPB5	-.037 ± .037	-.400 ± .130	-.040 ± .040	-
CRK5	-	.067 ± .381	-.400 ± .183	-
TOT5	-.161 ± .050	-.164 ± .157	-.129 ± .045	-
TOT6	-.193 ± .128	-.218 ± .142	-.169 ± .050	-
TOT7	-.213 ± .046	-.145 ± .125	-.149 ± .038	-.009 ± .009
TPA8	-.111 ± .055	-.429 ± .111	-.111 ± .055	-.143 ± .063
TPB8	-.097 ± .182	-.191 ± .170	-.136 ± .055	-.020 ± .020
CRK8	.012 ± .180	-.018 ± .176	.052 ± .191	-.049 ± .028
TOT8	-.033 ± .108	-.178 ± .101	-.034 ± .103	-.062 ± .021
TPA9	-.154 ± .054	-.292 ± .162	-.071 ± .036	-.091 ± .040
TPB9	-.184 ± .060	.237 ± .187	-.137 ± .051	-.018 ± .018
CRK9	-.478 ± .127	-.257 ± .223	-.091 ± .052	-
TOT9	-.226 ± .042	-.086 ± .112	-.100 ± .027	-.041 ± .017

TABLE 10.12

Mean F_{IS} for 3 loci, averaged over 3 mouse colonies

	Loci		
Census	*Hbb*	*Es-3*	*Es-1*
1	.185	-.408	-.091
2	.373	-.022	-.091
3	.566	.013	-.045
4	-.029	-.020	-.032
5	-.373	-.219	-.282
6	-.280	-.222	-.194
7	-.225	-.190	-.144
8	-.063	-.212	-.053
9	-.315	-.109	-.106
Grand mean	-.046	-.166	-.143

TABLE 10.13

Mean F_{IS} for 3 colonies, averaged over 3 loci

	Colonies		
Census	TPA	TPB	CRK
1	-.067	-.413	.056
2	-.083	.032	.235
3	-	-.103	.503
4	-.100	.107	-.091
5	-.353	-.306	-.153
6	-.368	-.123	-.176
7	-.269	-.081	-.208
8	-.282	-.144	.003
9	-.219	.032	-.201
Grand mean	-.284	-.110	-.016

standard errors for each F_{IS} are from Rasmussen (1964), but each sample represents a significant proportion of the colony from which it was collected, so the interpretation of these standard errors (which were derived assuming sampling with replacement) is not simple. The standard errors are provided so as to give a relative estimate of the precision of each estimate of F_{IS}.

As before, in Chapter 9, there is a wide range of F_{IS} values. The analysis of these data are complicated by two factors; firstly, the fact that each sample is a large proportion of each colony and secondly, the data from successive censuses are not independent of each other. Because of these two factors I will not attempt to carry out statistical tests on these data, but instead will calculate mean F_{IS} values in two different ways to bring out some points of interest. From equation 3.11, the mean fixation index in a set of n samples (ignoring sample sizes because interest centres upon the mean fixation index per sample, regardless of its size) is obtained from

$$F_{IS} = \sum_{i=1}^{n} q_i(1-q_i)F_{IS_i} \Big/ \sum_{i=1}^{n} q_i(1-q_i)$$

Table 10.12 presents the mean fixation index for each locus for each census and the grand mean over all censuses. The F_{IS} value for the *Hbb* locus is positive for censuses 1, 2 and 3 and reference to Table 10.11 shows that this is due to the positive values in CRK during these censuses. All other mean estimates of F_{IS} (except *Es-3* in census 3) are negative, as are the grand means. The grand mean for *Hbb* is less negative than for *Es-3* or *Es-1*, but this is clearly due to the positive values in the first three censuses. Table 10.13 gives the mean fixation index within each colony for each census and the grand mean over all censuses. The large positive F_{IS} values for *Hbb* in CRK for the first

three censuses are responsible for the positive mean F_{IS} values in CRK for that period. Most other means are negative.

The effect of the migrant male in the TPA colony who was heterozygous at the *Es-1* and *Es-3* loci and homozygous Hbb^d can be clearly seen in censuses 5 and 6 when the mean F_{IS} becomes very negative due to the number of heterozygous juveniles in the trappable population. The effects of this male are probably responsible for the more negative grand mean F_{IS} in TPA than the other colonies.

Some dramatic changes in mean F_{IS} values took place. For example, in colony TPB between censuses 1 and 2, and between censuses 4 and 5.

10.3.4 Approximate estimate of effective subpopulation size

The ecological and genetical data presented in this study indicate that all three colonies are distinct subpopulations. The negative F_{IS} values observed in Table 10.13 are evidence for the colonies being random mating demes (with immigration) similar to the mathematical random mating demes dealt with in Chapters 3 and 4. The important conclusion drawn from Chapters 3 and 4, that the expected value of F_{IS} in a random mating deme receiving immigrants is approximately $-(1-m)/N_e$, can be used here to estimate N_e. Providing that the mathematical theory is a reasonable approximation to the situation in these three mouse colonies, and if the 20% or thereabouts, of each colony that were missed at each census, as well as the effects of immigration on F_{IS}, are ignored, then the observed mean F_{IS} will yield a rough estimate of the average effective subpopulation size in these colonies. The mean F_{IS} over all censuses for the *Es-1*, *Es-3* and *Hbb* loci is -0.12. Equating this to $-\frac{1}{N_e}$, yields $N_e \doteqdot 8.3$. Because there are no standard errors attached to the mean F_{IS} and I do not know how the mathematical approximations that have been made will bias this estimate of N_e, little can be said about the confidence intervals surrounding this estimate.

However, this genetical estimate of N_e is remarkably close to the

TABLE 10.14 Population F_{ST} estimates, assuming sampling with replacement.

Census	*Hbb*	*Es-3*	*Es-1*	*Erp-1*
1. (a)	.234	.159	.019	-.005
(b)	.213	.150	.019	-.005
2. (a)	.213	.271	.019	-
(b)	.192	.243	.018	-
3. (a)	.320	.176	-.011	-
(b)	.281	.163	-.011	-
4. (a)	.563	.048	-.022	-
(b)	.465	.045	-.022	-
5. (a)	.222	.117	.092	-
(b)	.203	.110	.087	-
6. (a)	.077	-.024	.005	-
(b)	.074	-.024	.005	-
7. (a)	.036	.015	-.014	.017
(b)	.035	.014	-.014	.016
8. (a)	.011	-.015	-.012	.029
(b)	.011	-.015	-.012	.028
9. (a)	.044	-.014	-.007	.034
(b)	.043	-.014	-.007	.033

(a) s_q^2 from ordinary between samples component of variance.

(b) s_q^2 with Wright's correction applied.

TABLE 10.15 Population F_{ST} estimates, assuming 80% of each subpopulation sampled.

Census	*Hbb*	*Es-3*	*Es-1*	*Erp-1*
1. (a)	.254	.170	.037	.014
(b)	.233	.161	.037	.014
2. (a)	.254	.300	.051	-
(b)	.233	.271	.050	-
3. (a)	.354	.200	.014	-
(b)	.315	.187	.013	-
4. (a)	.588	.088	.016	-
(b)	.489	.085	.016	-
5. (a)	.241	.140	.118	-
(b)	.222	.134	.113	-
6. (a)	.099	.002	.031	-
(b)	.096	.001	.031	-
7. (a)	.053	.033	.005	.038
(b)	.052	.032	.005	.038
8. (a)	.026	-.002	.004	.043
(b)	.026	-.002	.004	.043
9. (a)	.055	.001	.007	.049
(b)	.054	.001	.007	.048

(a) s_q^2 from ordinary between samples component of variance.

(b) s_q^2 with Wright's correction applied.

ecological estimate obtained in 10.2.6, where the total effective population in TPA plus TPB was estimated as 13 (i.e. about 7½ in each) and the effective number in CRK was 9.6. Thus, it appears that genetical and ecological data can yield estimates of N_e that are very similar.

10.3.5 Genetic variation between colonies

The gene frequency data at the *A*, *Es-3* and *Hbb* loci in particular, show that the colonies differed markedly at the start of this study in July 1971, but they were more alike a year later. This observation can be quantified by estimating F_{ST} over the three colonies for each locus. I will use the "population" method rather than the "sample" method of Chapter 9 because the latter method appeared to yield larger values (i.e. may be an overestimate). There are two factors to be taken into account when calculating F_{ST} in this case. Firstly, should sampling with replacement (i.e. binomial sampling) or sampling without replacement (assuming 80% of each subpopulation taken at census) be assumed? The former assumption will subtract a larger sampling variance from the observed variance and is an attempt to estimate F_{ST} in the "population" from which the three colonies were drawn as samples. The latter assumption will be an attempt to estimate F_{ST} as it actually is in the population represented by the three colonies.

Secondly, the correction suggested by Wright (1943b) may or may not be applied (equation 9.8). There are thus four ways of estimating F_{ST}. The methods of calculation were the same as in Chapter 9 except that the weighted mean sampling error was different when it was assumed that 80% of each colony had been sampled. The i^{th} term in σ_S^2 in equation 9.6 was multiplied by $(1-0.8)(2N_i)/(2N_i-0.8)$ which is the appropriate correction for hypergeometric sampling when 80% of a population is sampled without replacement.

Tables 10.14 and 10.15 provide the four different estimates of F_{ST} for the *Erp-1*, *Es-1*, *Es-3* and *Hbb* loci. For the *Agouti* locus, equation 9.12 was used to estimate sample gene frequencies assuming $F_i = -0.12$.

TABLE 10.16

Population F_{ST} estimates for *Agouti* locus

Census	Assuming random sampling with replacement		Assuming 80% of each colony sampled	
	(a)	(b)	(a)	(b)
1	.467	.399	.483	.415
2	.529	.441	.553	.465
3	.335	.296	.357	.318
4	.763	.602	.776	.615
5	.608	.500	.622	.514
6	.565	.472	.575	.482
7	.193	.177	.220	.204
8	.100	.094	.125	.120
9	.376	.331	.384	.340

(a) S_q^2 from ordinary between samples component of variance.

(b) S_q^2 with Wright's correction applied.

TABLE 10.17

Mean estimate of F_{ST} over the *A*, *Es-1*, *Es-3* and *Hbb* loci.

Census	Assuming random sampling with replacement		Assuming 80% of each colony sampled	
	(a)	(b)	(a)	(b)
1	.220±.094*	.195±.079	.236±.093	.212±.079
2	.258±.105	.224±.087	.299±.103	.255±.085
3	.205±.080	.182±.071	.231±.081	.208±.072
4	.338±.193	.273±.154	.367±.186	.301±.150
5	.260±.119	.225±.095	.280±.117	.246±.092
6	.156±.138	.132±.115	.177±.134	.153±.112
7	.058±.046	.053±.043	.078±.048	.073±.045
8	.021±.027	.020±.025	.038±.030	.037±.028
9	.100±.093	.088±.082	.112±.092	.101±.081

(a) S_q^2 from ordinary between samples compenent of variance.

(b) S_q^2 with Wright's correction applied.

* Standard error of mean.

(see Table 10.6). This procedure markedly reduced the values of F_{ST} obtained in comparison with those obtained when $F_i = 0$ in these calculations. The correction used for calculating σ_i^2 (equations 9.11 and 9.16) for sampling without replacement was $(1-0.8)(N_i)/(N_i-0.8)$. The estimates of F_{ST} for the *Agouti* locus are given in Table 10.16. For each census, F_{ST} is much larger at the *Agouti* locus than any of the others because of the marked difference in phenotype frequencies in TPA and TPB.

The mean estimates of F_{ST} for all loci except *Erp-1* (which was usually monomorphic) are given in Table 10.17. F_{ST} was large in the winter and spring of 1971, rose even higher at the time of the population crash in CRK, and descended rapidly in the summer of 1972. There was virtually no significant differentiation of gene frequencies by autumn 1972 (censuses 7 and 8).

The initially high levels of F_{ST} can be explained by postulating that only very few mice survived the population crash following the mouse plague in 1970 (the crash took place about a year before this study began). These few were the founders of the colonies and at the time I began this study, the "founder effect", followed by perhaps two generations of inbreeding, had led to marked genetic differentiation of the colonies. But considerable amounts of gene flow into the colonies during the year of my study then eliminated the genetic differentiation that had built up. There is considerable evidence for gene flow into the three colonies. The mark-recapture experiments showed that three mice made movements between colonies. The genetic data from the TPA colony showed that at least two immigrants entered that colony. After the population crach in the CRK colony, there were insufficient mice left to re-establish this colony on their own, so there must have been much immigration into this colony. The entry of $Erp\text{-}1^b$ alleles into all colonies late in the study shows that considerable movement of mice into the colonies must have been taking place.

I had hoped to be able to use the regression of Δq on q (Tamarin and Krebs, 1969, 1973) to make quantitative estimates of the amount of gene flow into each colony. However, as shown in Appendix 3, this procedure is too biassed to be of any use in this case.

10.3.6 Old Field study

The aim of this study was to find genetical evidence of population subdivision within the OF area. The genotype of each mouse was plotted on a map at the site where it was first caught. Visual inspection of the data showed no evidence of genetic heterogeneity between different parts of the old field: the genotypes appeared to be distributed randomly. The area was divided into four quarters and the gene frequencies were tested for homogeneity at the *Es-3* and *Hbb* loci. The χ^2_3 values were not significant ($\chi^2_3 = 0.604$ for *Es-3*, $\chi^2_3 = 5.94$ for *Hbb*). The numbers were too small at the *A*, *Es-1* and *Erp-1* loci for χ^2_3 tests to be made, but as mentioned before, there was no evidence for clustering of similar genotypes.

This population had only recently expanded and so there were probably many immigrants from other areas as well as juveniles born on the OF area all intermingled. The movement of mice trying to find fresh cracks in the soil and dig burrows would tend to prevent population subdivision within the old field.

The results of Selander (1970) where he found genetic evidence of subdivisions within chicken barns are not confirmed. Probably the fact that OF is a transient population (reduced to very low numbers or extinction each winter) prevents the development of subdivision by territoriality and parent-offspring interbreeding. The high immigration rates observed in the other three colonies would also prevent population subdivision occurring in the old field.

10.4 DISCUSSION

The ecological data collected in this study show that even in the presence of a food surplus, the size of a mouse population is severely limited by the amount of usable living space (in this case, burrows). During the spring and summer months there is an export of juvenile mice from established mouse colonies. By the autumn, most of the older mice who have been breeding since spring have died and in the autumn a cohort of juveniles remains in the colony and over the next winter (Anderson, 1970).

The number of adult mice under each feedstation was remarkably small (usually two or three), and this might be taken as evidence for territoriality. The limited mobility of mice between feedstations would also support the territoriality hypothesis. However, the hypothesis that mice defend territories against foreigners and actively drive out intruders is at variance with the observed gene flow into these colonies. The data support the view that juveniles are driven out (or wander off of their own accord) but do not support the hypothesis that gene flow between mouse subpopulations is prevented by aggression against strangers. How else could the *Es-1*^{a}, *Es-3*^{b} and *Hbb*^{d} alleles have got into the TPA colony, or the *Erp-1*^{b} alleles into all colonies, if strange mice were not accepted? Experiments with confined mice (e.g. Oakeshott, 1974) suggest that female immigrants are more readily accepted into confined colonies than males. But the observation that juveniles of distinctive genotypes were produced under two feedstations simultaneously shows that at least one male was able to enter TPA and become reproductively dominant in two of the three feedstations. These are the first field observations of gene flow into mouse subpopulations. They show that for this population, at least, territoriality is probably not a great hindrance to gene flow.

The small effective population size of mouse subpopulations appears to be confirmed by this study. But it should be noted that the size of each colony was limited by the amount of food and shelter available.

The complete destruction of the CRK colony in early summer 1971, and its partial destruction by flooding in September 1971, and by rats in July 1972, serves to show how vulnerable is a mouse colony. Predators, climatic disasters, food shortages, etc. can terminate the "life" of a colony, killing the mice or sending them in search of new habitable places.

The mean gene or phenotype frequencies at all loci had returned by July 1972, to the values estimated in the height of the mouse plague two years earlier. This observation does not support the hypothesis that different optimum gene frequencies are selected in mouse populations of different densities (e.g. Tamarin and Krebs, 1969). The mean gene frequencies could be the same in 1970 and 1972 because of similar selection pressures or (if the genes were neutral) simply the lack of random sampling drift in the total population of all mice living in haystacks, barns, houses, field, etc. within many kilometres of Turretfield. Random sampling drift would occur in small colonies left after the plague, but when numbers increased in 1971-72 and gene flow eliminated the isolation of these colonies, the whole population would still have almost the same gene frequencies as before.

CHAPTER 11

CONCLUSIONS

The theory of inbreeding in subdivided populations has been well studied in terms of F_{IT}, F_{ST}, $\bar{f}$ and f_o. But the mathematical theory of F_{IS} has been neglected until now. Chapters 3 and 4, where many formulae useful in the study of populations with non-overlapping generations have been derived, represent a start on the theory of F_{IS}.

The apparent gap between the F-statistics and the probability of identity methods was bridged in Chapter 5 after a suitable definition of F_{ST} at a multi-allelic locus was obtained. The recent work of Cockerham (1973) has improved understanding of F-statistics (although Cockerham prefers to use a different terminology) and paid more attention to the problems of estimating the different F-statistics (unfortunately, this paper became available too late for Cockerham's methods to be used in this thesis). The F-statistics provide a useful tool in the analysis of data from subdivided populations in nature or domestication. The only previous attempt to use F-statistics in a complete analysis was Neel and Ward (1972), but they unfortunately used Nei's (1965) definition of the mean F_{IS} and they were confused by their results.

The theoretical work reviewed in Chapter 2 suggested that population subdivision would not have much effect on the evolutionary behaviour of neutral alleles unless the number of migrants per subpopulation was less than one. The observed genetic differentiation of mouse subpopulations in the City of Adelaide and the State of South Australia is sufficiently large ($F_{ST} \doteqdot 25\%$) that the numbers of neutral alleles per locus and the rates of change of frequency of neutral alleles would be markedly affected by population subdivision.

In the theoretical review, it was also suggested that the causes of genetic differences between subpopulations would be difficult to identify with precision. The comparison of ecological and genetical data on four mouse populations revealed that the amount of genetic differentiation was inversely proportional to the estimated number of immigrants per subpopulation (Nm). This demonstrated the sizes of subpopulations (N) and amounts of immigration into subpopulations (m) are important factors in controlling the differentiation of subpopulations. But no evidence could be obtained about the role that selection played in increasing or decreasing the genetic variation between subpopulations.

When the regression between genetic and geographic distances between subpopulations within each of six populations was studied, the results did not agree with any existing theory. Geographic distance did not appear to have much influence on genetic distances until the geographic distances exceeded about 120 km. Mathematical theory based on neutral alleles suggests that genetic distances should increase with geographic distances initially but should be effectively independent of geographic distances at greater distances. This prediction is the opposite of observation, and is clearly wrong. No wholly suitable explanation has been proposed for the difference between theory and observation. Long range gene flow or selection for genetic equilibria that gradually alter across a landscape have been suggested, but there are no rigorous mathematical analyses of alternative models to help decide which, if either, explanation is adequate.

The trap-recapture study is the first of its nature. It was intended that estimates of effective subpopulation sizes and immigration rates would be obtained from ecological and genetical data. The first objective was achieved: both ecological and genetical data suggest that the average effective breeding size of the colonies was a little less than ten. The estimation of immmigration rates was only partially successful. From

capture-release-recapture data, it was clear that female mice could travel between colonies, and at a rate perhaps exceeding 5%. The genetical data revealed that many genes were entering each colony, and in one fortunate case, genetic data showed that a foreign male had entered a colony and become reproductively dominant over 2/3rds of it. Observed levels of inbreeding due to population subdivision plummeted during the study period, providing evidence of high rates of immigration following a period when small colony sizes and negligible immigration rates had allowed the colonies to undergo marked random sampling drift. The population crash following the mouse plague of the previous year was thought to be the cause of the low numbers and limited dispersal before the study began.

Mouse populations living in different environments have different ecological and genetical structures. For this reason it is impossible to generalise about the significance of population subdivision in the evolution of mouse populations. In the grain-growing areas north of Adelaide, levels of population subdivision are low (probably higher than in the 1970 plague and more like the situation at Turretfield in 1972).

In the City of Adelaide and (by comparison with the data of Petras *et al.*, 1969) in Ontario, population subdivision probably is important in the evolution of these populations.

Until more studies of population subdivision are carried out in a variety of species of organisms, little can be written confidently about the importance of population subdivision generally. The work done in this thesis shows that the theory of population subdivision is poorly developed and that there is a lack of detailed studies on natural populations. Much research remains to be done in this field of ecological genetics.

APPENDIX 1

THE GENETICS OF AN ELECTROPHORETIC VARIANT OF AN ERYTHROCYTIC PROTEIN IN THE HOUSE MOUSE (*Mus musculus*)

Submitted to *"Animal Blood Groups and Biochemical Genetics"*;
July 1974.

The genetics of an electrophoretic variant of an erythrocytic protein in the house mouse (*Mus musculus*)

Running title: Erythrocytic protein variation in mouse

Author: G.C. Kirby

Address: Department of Genetics,
University of Adelaide,
South Australia 5001.

Summary

An electrophoretic mobility variant of a protein band anodal to haemoglobin in alkaline starch gels was found in wild house mouse populations in South Australia. Most wild mice and all inbred lines examined are homozygous for the *Erp-1*^{a} allele at the locus controlling this variation. The rare allele, *Erp-1*^{b}, which has a frequency of about 1% in South Australia, produces a protein band of slower mobility in alkaline gels. Homozygotes show one protein band whilst the heterozygote has three bands. The *Erp-1* protein does not appear to be haemoglobin or carbonic anhydrase. The *Erp-1* locus is closely linked to *Es-1* on chromosome 8 with a recombination fraction of 6.04 ± 1.32%.

Introduction

Electrophoretic variation at several loci can be identified in house mouse (*Mus musculus*) haemolysates by gel electrophoresis. The product of most of these loci can only be detected on a gel by using specific staining techniques that utilise the enzymatic functions of these proteins. Only two systems showing genetically controlled variation have been reported in which sufficient protein exists in haemolysates for it to be detectable with a general protein stain such as Amido Black: *Hbb* (haemoglobin β-chain) and *Car-1* (was *Pro-1* (Biddle and Petras, 1967) but name was changed because it was a carbonic anhydrase (Biddle and Krasny, 1970; Biddle, 1974)). This note records the discovery of a third such system.

When mouse haemolysates are electrophoresed in alkaline gels a protein band is normally found anodal to the haemoglobin zone (if there is any serum contamination of the haemolysate, then under these conditions this band will be cathodal to the albumin band). An electrophoretic mobility variant of this protein was found in wild mouse populations in South Australia and two mice carrying the variant were mated to laboratory stocks to enable its inheritance to be studied. Until the function of the protein can be elucidated, the locus determining the observed variation will be denoted by *Erp-1*, for erythrocytic protein.

Materials and methods

Blood samples from the suborbital sinus were collected in a heparinised syringe, centrifuged and the erythrocytes washed three times in 0.87% NaCl. After the final wash the packed erythrocytes were frozen at $-30^{\circ}C$ for at least several days, thawed, mixed with an equal volume of distilled water and refrozen. Before electrophoresis the samples were thawed and centrifuged at 1,500g for 30 minutes.

Initially, Smithies (1959b) gel buffer (0.045M Tris, 0.0008M Na_2EDTA,

0.025M H_3BO_3, pH 8.6) was used, but with this buffer the protein band of the variant homozygote was overlaid by haemoglobin and could not be seen. In order to slow the haemoglobin relative to the variant *Erp-1* protein so that they became separated on a gel, the buffer was modified by adding sufficient Boric acid to lower the pH to about 7.5. Electrolyte buffer was a one-third dilution of gel buffer. Starch gels (13% Connaught starch) were run horizontally in a refrigerator with 10-15 volts/cm for up to 2½ hours. Proteins were detected on gels by staining with Amido Black. Two methods of detecting peroxidase activity were used, the o-dianisidine method (Owen and Smith, 1961) and the benzidine method (Smithies, 1959a). Carbonic anhydrase was stained for by the bromothymol blue-CO_2 method (Pihar, 1968; Biddle, pers. comm.), whilst esterases and phosphatases were detected using both the naphthyl salts and the 4-methyl-umbelliferyl salts of acetate and phosphate respectively (cf. Hopkinson *et al.*, 1973).

Results

All inbred lines studied (Balb/c, CBA, C3H, C57B1, DBA, SWR) and most wild mice had a single protein band anodal to the haemoglobin when Smithies' buffer was used. A few wild mice had two protein bands; one with the same mobility as the commonly occurring band (although now fainter than usual) and the other nearer the haemoglobin zone. The common phenotype was designated Erp-1A and the variant Erp-1AB. When two Erp-1AB animals were mated together, about one-quarter of the progeny (denoted Erp-1B) did not appear to have any protein bands anodal to the haemoglobin when run on Smithies' buffer. When the modified pH 7.5 buffer was used it was found that the Erp-1AB mice actually had three protein bands and the Erp-1B mice had a single protein band of the same mobility as the slowest (most cathodal) band in the Erp-1AB mice (see Fig. 1).

Table 1

The inheritance of *Erp-1* phenotypes. The χ^2 tests are of the hypothesis that there are two codominant alleles at one locus.

Cross ♀	x	♂	Number of matings	Progeny phenotype Erp-1A	Erp-1AB	Erp-1B	χ^2
Erp-1A	x	Erp-1A	8	92	0	0	
Erp-1B	x	Erp-1B	2	0	0	17	
Erp-1A	x	Erp-1B	4	0	21	0	
Erp-1B	x	Erp-1A	3	0	17	0	
Erp-1A	x	Erp-1AB	10	77	62	0	1.62(1:1)
Erp-1AB	x	Erp-1A	9	66	63	0	0.07(1:1)
Erp-1AB	x	Erp-1AB	7	33	74	31	0.78(1:2:1)

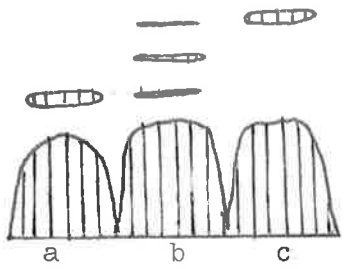


Fig. 1. Electrophoretic patterns of *Erp-1* genotypes in Tris-EDTA-Borate buffer at pH 7.5, (a) variant homozygote (*Erp-1*^{b}*Erp-1*^{b}), (b) heterozygote (*Erp-1*^{a}*Erp-1*^{b}), (c) normal homozygote (*Erp-1*^{a}*Erp-1*^{a}).

Data on the inheritance of the *Erp-1* phenotypes are given in Table 1. These data are consistent with the hypothesis that the variation at the *Erp-1* locus is due to two codominant alleles, denoted $Erp\text{-}1^a$ and $Erp\text{-}1^b$. The phenotypes Erp-1A, Erp-1AB and Erp-1B are produced by the genotypes $Erp\text{-}1^aErp\text{-}1^a$, $Erp\text{-}1^aErp\text{-}1^b$ and $Erp\text{-}1^bErp\text{-}1^b$ respectively.

No linkage between *Erp-1* and the agouti, albino, *Hbb* or *Es-3* locus was found, but crosses involving *Es-1* showed close linkage. From backcrosses using double heterozygotes for *Es-1* and *Erp-1*, there were 7 recombinants out of 79 progeny from female heterozygotes and 3 recombinants out of 95 progeny from males. The difference in recombination frequency between the two sexes is not statistically significant and the pooled data yield a recombination fraction (r) of 5.7 ± 1.8%. Matings of double heterozygotes in the same linkage phase (both parents in coupling or both in repulsion) yielded 81 F_2 progeny from which r was estimated by maximum likelihood (cf. Robinson, 1971) as 6.4 ± 1.99%. The joint estimate of r from backcross and F_2 data was 6.04 ± 1.32% by maximum likelihood iteration.

Attempts are being made to obtain further linkage data with loci linked to *Es-1*.

The $Erp\text{-}1^b$ gene is rare but widely scattered in mouse populations in South Australia. Of 840 wild mice caught in diverse regions of the State, 18 were heterozygotes. No Erp-1B mouse was found.

Attempts to show that the *Erp-1* protein had peroxidase, carbonic anhydrase, esterase or phosphatase activity all failed.

Discussion

The linkage between *Erp-1* and *Es-1* shows that *Erp-1* is located on chromosome 8 (linkage group XVIII) near to, or within, a dense cluster of loci. This region of about 11 recombination units in chromosome 8 carries five enzyme loci: *Es-1* (Popp and Popp, 1962); *Es-2* (Petras, 1963); *Es-5* (Petras and Biddle, 1967); *Es-6* (Petras and Sinclair, 1969); *Got-2*

(DeLorenzo and Ruddle, 1970); and three loci controlling behavioural and morphological variants: tottering (*tg*), Oligosyndactylism (*Os*) and amputated (*am*) (Green, 1966).

Including *Erp-1*, there are now four loci (possibly five if a new esterase variant, *Es-7*, turns out to be separate from *Es-2*; Chapman, pers. comm.) known to be on a relatively short region of chromosome 8 and all possessing codominant alleles. Thus this region could prove very useful for studies of recombination events in mammalian chromosomes or for studies of linkage disequilibrium in natural or artificially established mouse populations. A useful feature of these loci is that they can all be scored in only two tissues: erythrocytes (*Erp-1* and *Es-7*, if confirmed) and kidney (*Es-1*, *Es-2* and *Got-2*).

The function of the *Erp-1* protein is unknown. It is presumably in high concentration in the erythrocytes for it stains with Amido Black. The proteins in highest concentration in erythrocytes are haemoglobin and carbonic anhydrase but without detectable peroxidase or carbonic anhydrase activity, the *Erp-1* protein is unlikely to have either function. Human carbonic anhydrase is a monomer and haemoglobin, despite its biquaternary structure, does not usually show hybrid bands in heterozygotes except in fish and two unusual human variants (Manwell and Baker, 1970). The observation that the $Erp\text{-}1^{a}Erp\text{-}1^{b}$ heterozygote shows a hybrid band indicates that the *Erp-1* protein is at least a dimer, and also suggests that the *Erp-1* protein is not carbonic anhydrase or haemoglobin. It might be anticipated that proteins concerned with the structure of erythrocytes and their membranes could occur in high concentration in haemolysates. Thus the possibility exists that the *Erp-1* protein is involved in the structure of the erythrocyte.

The *Erp-1* locus is different from the *Car-1* locus because *Car-1* is not linked to *Es-1*, has carbonic anhydrase activity and heterozygotes do not show a hybrid protein band.

Acknowledgements

I am grateful to Professor J.H. Bennett, Dr. R.M. Hope and Mr. J. Oakeshott for many improvements to the manuscript. Mr. C. Chesson provided mouse stocks and much assistance and Mr. I.R. Goodwins obtained the photograph for Fig. 1. Dr. A.G. Searle kindly suggested the name, *Erp-1*.

References

Biddle, F.G., 1974. *Mouse News Letter* No. 50: 3.

Biddle, F.G. & H. Krasny, 1970. The association of carbonic anhydrase activity with the non-haemoglobin erythrocytic protein (*Pro-1*) of the house mouse. *Can. J. Genet. Cytol.* 12: 374.

Biddle, F.G. & M.L. Petras, 1967. The inheritance of a nonhaemoglobin erythrocytic protein in *Mus musculus*. *Genetics* 57: 943-949.

DeLorenzo, R.J. & F.H. Ruddle, 1970. Glutamate oxalate transaminase (*Got*) genetics in *Mus musculus*: linkage, polymorphism and phenotypes of the *Got-2* and *Got-1* loci. *Biochem. Genet.* 4: 259-273.

Green, M.C., 1966. Mutant genes and linkages. In *Biology of the Laboratory Mouse*, 2nd edition. Ed. E.L. Green. McGraw-Hill, New York.

Hopkinson, D.A., M.A. Mestriner, J. Cortner & H. Harris, 1973. Esterase D: a new human polymorphism. *Ann. Hum. Genet. Lond.* 37: 119-137.

Manwell, C. & C.M.A. Baker, 1970. *Molecular Biology and the Origin of Species*. Sidgwick & Jackson, London.

Owen, J.A. & H. Smith, 1961. Detection of ceruloplasmin after zone electrophoresis. *Clin. Chim. Acta* 6: 441-444.

Petras, M.L., 1963. Genetic control of a serum esterase component in *Mus musculus*. *Proc. Nat. Acad. Sci.* 50: 112-116.

Petras, M.L. & F.G. Biddle, 1967. Serum esterases in the house mouse, *Mus musculus*. *Can. J. Genet. Cytol.* 9: 704-710.

Petras, M.L. & P. Sinclair, 1969. Another esterase variant in the kidney of the house mouse, *Mus musculus*. *Can. J. Genet. Cytol.* 11: 97-102.

Pihar, O., 1968. Localization of carbonate hydro-lyase activity in starch gel electrophoresis. *Z. Klin. Chem. u. Klin. Biochem.* 2: 110-111.

Popp, R.A. & D.M. Popp, 1962. Inheritance of serum esterases having different electrophoretic patterns among inbred strains of mice. *J. Hered.* 53: 111-114.

Robinson, R., 1971. *Gene Mapping in Laboratory Mammals*. Part A. Plenum Press, London.

Smithies, O., 1959a. Zone electrophoresis in starch gels and its application to studies of serum proteins. In *Advances in Protein Chemistry*, Vol. 14 (eds.) Anfinson, C.B., Anson, M.L., Bailey, K. & Edsall, J.T. p.65. Academic Press, New York.

Smithies, O., 1959b. An improved procedure for starch gel electrophoresis: further variations in the serum proteins of normal individuals. *Biochem. J.* 71: 585.

APPENDIX 2

GREYING WITH AGE: A COAT-COLOR VARIANT IN WILD AUSTRALIAN POPULATIONS OF MICE

Published in: *"Journal of Heredity"* Volume 65; pp.126-128,1974.

GREYING WITH AGE: A COAT-COLOUR VARIANT IN WILD AUSTRALIAN POPULATIONS OF MICE

Most mammals (except humans) retain full hair pigmentation up to the end of their lifespan, although a few species are known to have genes that cause premature greying of the pelage. Searle[5] reviewed the genes known to cause age-related greying in the following animals: black and brown rats, *Peromyscus maniculatus*, guinea-pigs, rabbits, dogs, sheep, and horses. The house mouse (*Mus musculus*) is notably absent from this list, although Gruneberg[2] mentions two cases of mice developing white hairs after reaching adulthood, and silver (*si*) homozygotes sometimes do not develop unpigmented hairs until maturity[1].

This note reports the discovery of mice in wild populations in South Australia with a grey appearance (due to unpigmented hairs) that intensifies with ageing. The first cases were found in 1970 during a mouse plague in wheatgrowing areas on Yorke Peninsula. Three were caught near Price in a sample of 299 mice, another two near South Kilkerran and one at Alford. (South Kilkerran is 43 miles from Alford, with Price inbetween.) At South Kilkerran, Alford and two other localities, several hundred mice were quickly surveyed for morphological variants in groups of 30 or more in a large bin. Another mouse was obtained in 1971 from a suburb of Adelaide (about 80 miles overland from the other localities) that was initially phenotypically normal but began turning grey after a month. An examination of mouse skins in the South Australian Museum (kindly provided by Mr. P. Aitken) turned up a specimen caught at Kadina on Yorke Peninsula in 1953 (S.A. Museum registration number M5946) that closely resembled the grey mice mentioned above and their progeny.

TABLE I. Results of F_1 x F_1 crosses (line 1) and testcrosses of grey progeny of grey x grey matings (lines 2 and 3).

	♀ x ♂	No. of matings	Progeny phenotypes*		χ^2_1
			Grey	Normal	
(1)	*Gaga* x *Gaga*	8	33	13	0.26(3:1)
(2) C57BL	(*gaga*) x *Ga.*	5	7	65	46.5 (1:1)
(3)	*Ga.* x C57BL (*gaga*)	3	12	0	12.0 (1:1)

* Scored at 10-12 months of age.

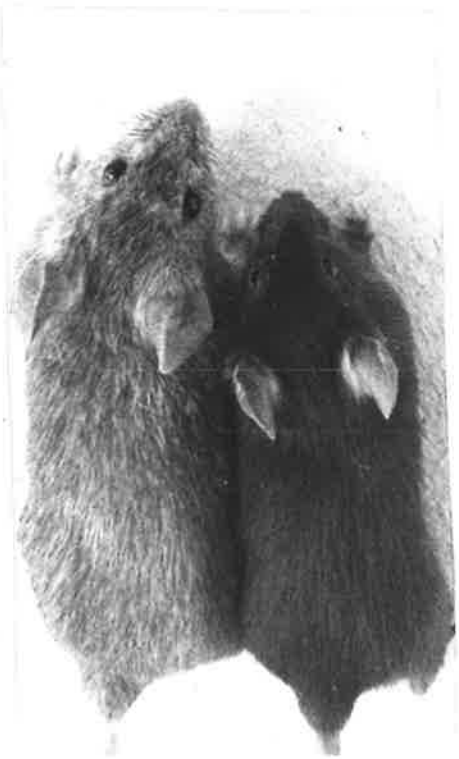


FIGURE 1 Female 181 (age 8½ months, genotype *aaBBGa*.) showing grey phenotype of *Ga* alongside a 5-week-old progeny that shows normal black (*aaBB*) phenotype.

DESCRIPTION

One of the six grey mice from Yorke Peninsula and the one from Adelaide bred in the laboratory. The stocks derived from these two mice have been kept separate so it is not certain that the same gene is involved in each case. However, the grey phenotype is essentially the same in both stocks, and tentatively, it is considered to be controlled by the same gene in each, denoted *Ga* for greying with age.

Table I presents some data on the inheritance of *Ga*. The first line (and other data) demonstrate an autosomal dominant mode of inheritance. The second and third lines show the results of crossing the grey progeny of two grey parents to C57BL (genotypically *aaBB*) mice. The progeny of these matings were kept for 10-12 months or until they became grey. Two comments can be made: (1) when the female parent was C57BL, the proportion of grey mice in the progeny was much lower than when the male parent was C57BL ($\chi^2_1 = 48$). The seven grey progeny of C57BL females were only detected after they were more than 10 months old, suggesting that more progeny would have become grey if they were kept for a longer time. All the progeny of *Ga* females became grey in less than 10 months. There is clearly a strong maternal effect upon the time taken to become grey. (2) The fact that all 12 progeny of *Ga* females turned grey makes it unlikely that all of the females were heterozygous, i.e. one or more were probably homozygous *GaGa*, evidence that *GaGa* mice are viable and fertile.

A comparison of the general appearance of a mature grey mouse (genetically *aaBBGa*) with that of an immature black progeny is shown in Figure 1. Loss of hair pigmentation becomes apparent at ages varying from 3 months (after three generations of selection for early greying) to over 12 months (after crossing to C57BL mice). After breeding the *Ga* gene into an *aaBB* (all black) stock it was observed

that hairs on the belly lose pigment earliest, then usually the face and hindquarters, with the dorsal midline being the last area to remain black. There is variable expression of the gene in old mice, some being only faintly grey while others are nearly white. Microscopic examination of hairs from *Ga* mice showed most hairs to be either fully pigmented or completely lacking in melanin granules and only rarely were hairs partially pigmented.

There is a marked resemblance between these mice and the silver (*sisi*) mice illustrated in Dunn and Thigpen[1]. No crosses have been made to prove that *Ga* and *si* are nonallelic. The *Ga* gene differs from *si* in that it is dominant, few hairs are a mosaic of pigmented and unpigmented regions, and greying does not begin until animals are several months old, becoming progressively more pronounced with advancing age. In silver mice on certain genetic backgrounds, the amount of silvering decreases with age.

DISCUSSION

The discovery of *Ga* is worthy of record for the following reasons. This is an example of a morphological variant, which is present in widespread local populations of the house mouse in South Australia, that may have been present for at least 20 years. The exact frequency of grey mice in wild populations is hard to establish because most *Ga* mice in the laboratory do not express the gene until about 6 months of age and relatively few wild mice reach that age (Newsome[3,4], Kirby, unpublished). Hence, most wild mice are too young to express this character. This gene puts *Mus musculus* into the list of animals known to have genes causing progressive greying with age (cf. Searle[5]). Since the expression of *Ga* is age-dependent, it may be useful in studies of ageing and especially the effects of various environments on ageing (O. Mayo, personal communication).

SUMMARY

Greying with age (*Ga*), an autosomal dominant gene that causes progressive loss of hair pigmentation with ageing in mature house mice, was found in wild mice from several localities in South Australia. The time involved in becoming grey is influenced by maternal effects.

Acknowledgements: The author gratefully acknowledges the assistance of Professor J.H. Bennett and Dr. O. Mayo, who gave useful advice on a draft manuscript; Mr. C. Chesson, who assisted with mouse stocks; and Mr. J.R.B. Walpole, who helped with photography.

REFERENCES

1. DUNN, L.C. and L.W. THIGPEN. The silver mouse, a recessive color variation. *J. Hered.* 21:495-498. 1930.

2. GRUNEBERG, H. The Genetics of the Mouse. Nijhoff. The Hague, 1952.

3. NEWSOME, A.E. A population study of house-mice temporarily inhabiting a South Australian wheatfield. *J. Anim. Ecol.* 38:341-359, 1969.

4. ————————. A population study of house-mice permanently inhabiting a reed-bed in South Australia. *J. Anim. Ecol.* 38:361-377, 1969.

5. SEARLE, A.G. Comparative Genetics of Coat Colour in Mammals. Academic Press/Logos Press. New York/London, 1968.

APPENDIX 3

THE BIAS IN THE REGRESSION OF Δq ON q

Published in: *"Heredity"* August,1974.

THE BIAS IN THE REGRESSION OF Δq ON q

Department of Genetics, University of Adelaide, Adelaide, South Australia

SUMMARY

It is shown that when the regression of Δq (the change in gene frequency) on q (gene frequency) is calculated from data obtained from a sequence of censuses of one population, the regression coefficient does not provide an unbiased estimate of the amount of immigration or selection maintaining an equilibrium gene frequency in the population. Also, the equilibrium gene frequency estimated from the regression will be biased towards the observed gene frequency in the first census.

The relationship between the gene frequency in a population at time t (q_t) and the change in gene frequency (Δq_t) over the next time interval is useful for locating gene frequency equilibria in theoretical models of population genetics (Li, 1955). At an equilibrium point, $\Delta q_t = 0$. If there is a stable equilibrium gene frequency ($\bar{q}$) maintained by recurrent mutation, frequency dependent selection or immigration from a source population with constant gene frequency, then the expected change in gene frequency over one generation is

$$\Delta q_t = x(\bar{q} - q_t) \qquad \ldots\ldots (1)$$

where x is a constant which is a measure of the "force" by which the gene frequency is returned to $\bar{q}$ after perturbation. If the equilibrium is maintained by heterosis, then a similar equation will be valid except

that x will be approximately constant only when q_t is near $\bar{q}$ (Kimura & Weiss, 1964). The simplicity of this approach to the determination of a gene frequency equilibrium and the "force" maintaining the equilibrium in theoretical models suggests that a similar approach may perhaps be useful when examining data from natural populations with the intention of demonstrating selection pressures or gene flow. When q_t has been estimated in a population on many successive censuses then the regression of Δq_t on q_t can be calculated to obtain an equation similar to equation (1). The slope of the regression, b, will be an estimate of -x, and an estimate of $\bar{q}$ is obtained by solving the regression equation for $\Delta q = 0$. This regression procedure has been noted by Williamson (1972) (although he plots $\Delta q/q(1-q)$ against q, which is more preferable if heterosis is suspected) and applied by Tamarin and Krebs (1969, 1973) to *Microtus* population data.

When a modification of this method was applied to data from three house-mouse populations that were studied by mark-recapture techniques for a year, calculated values of b yielded estimates of x that were too large (0.48 to 1.3) to be accepted (Kirby, unpub.). This observation suggested that the regression might be biased so as to produce estimates of x that are much too large and it is easy to show that this is the case. Consider a locus with two neutral alleles in a small population which receives immigrants at a steady rate m per generation from a source population with constant gene frequency $\bar{q}$. Then the gene frequency in generation t+1(q_{t+1}) will be given by

$$q_{t+1} = q_t + m(\bar{q} - q_t) + \delta q_t \qquad \ldots\ldots (2)$$

where δq_t is the change in gene frequency due to random drift.

If the regression of Δq_t on q_t is calculated, the slope will be

$$b = \frac{\mathrm{cov}(q_t, \Delta q_t)}{\mathrm{var}(q_t)}$$

$$= \frac{\mathrm{cov}(q_t, q_{t+1})}{\mathrm{var}(q_t)} - 1 \qquad \ldots\ldots (3)$$

$$\doteq r(q_t, q_{t+1}) - 1 \qquad \ldots\ldots (4)$$

where $\mathrm{cov}(q_t, \Delta q_t)$ is the covariance of q_t and Δq_t, $\mathrm{var}(q_t)$ is the variance of q_t and $r(q_t, q_{t+1})$ is the correlation between q_t and q_{t+1}. From equation (4), the value of b is approximately limited to the range 0 to -2, that is, the estimates of m (derived from -b) will be restricted to the range 0 to 2. When there is no correlation between q_t and q_{t+1}, m will be +1, but in data collected from natural populations a positive correlation between successive gene frequencies would be expected and m will usually be less than 1.0. If equation (2) is substituted into equation (3), it is possible to obtain

$$b = -m + \frac{\mathrm{cov}(q_t, \delta q_t)}{\mathrm{var}(q_t)} \qquad \ldots\ldots (5)$$

Thus b will decrease with increasing m, and the value of b may also be affected by the size of the deme (as $E(\delta q_t)^2$ is proportional to 1/(deme size)) and the length of time over which the censuses are made, as $\mathrm{var}(q_t)$ will increase with time. The second term in the R.H.S. in equation (5) represents the bias in the estimation of -m by b. As $(\delta q_t)^2$ is proportional to $q_t(1-q_t)$, $\mathrm{cov}(q_t, \delta q_t)$ will not usually be zero.

For a given population, it may be possible to estimate the magnitude of this bias, $\mathrm{cov}(q_t, \delta q_t)/\mathrm{var}(q_t)$, and hence to compensate for it, if the bias is influenced in a simple manner by only a few

TABLE 1

Estimation of the coefficient of the regression of Δq on q for various initial gene frequencies and immigration rates.

q_o	m	NGENS		
		10	20	40
.1	0	-.4508 ± .0294	-.2357 ± .0235	-.1981 ± .0137
	.1	-.2434 ± .0185	-.1822 ± .0112	-.1391 ± .0061
	.2	-.2679 ± .0131	-.2544 ± .0083	-.2212 ± .0067
.2	0	-.4114 ± .0329	-.2398 ± .0213	-.1147 ± .0093
	.1	-.3508 ± .0272	-.2107 ± .0146	-.1487 ± .0081
	.2	-.3514 ± .0209	-.2572 ± .0104	-.2443 ± .0079
.3	0	-.3702 ± .0301	-.2179 ± .0195	-.1220 ± .0097
	.1	-.3825 ± .0305	-.2310 ± .0150	-.1921 ± .0103
	.2	-.4164 ± .0242	-.3034 ± .0142	-.2734 ± .0106
.4	0	-.3980 ± .0318	-.1831 ± .0148	-.1222 ± .0085
	.1	-.4660 ± .0279	-.2712 ± .0168	-.1917 ± .0090
	.2	-.5404 ± .0305	-.3432 ± .0190	-.2810 ± .0114
.5	0	-.4195 ± .0284	-.2185 ± .0179	-.1217 ± .0101
	.1	-.4244 ± .0283	-.3095 ± .0193	-.2147 ± .0112
	.2	-.5631 ± .0329	-.3911 ± .0208	-.2761 ± .0106

q_o = initial gene frequency, m= immigration rate from population with gene frequency of 0.5; NGENS = number of generations over which the regression is calculated. For each estimate, 100 replicate simulations were made and the standard error of each estimate is given.

parameters. With this possibility in mind, a number of computer simulations were made of a haploid population with non-overlapping generations. Immigrants entered the population at rate m per generation, where m is the probability of a gene in the next generation coming from a source population with gene frequency 0.5 instead of from the genes in the present generation (sampling with replacement). In the first set of simulations three population sizes (100, 200 and 400 genes) and three immigration rates (0, 10% and 20%) were used and simulations were run for three different time periods (10, 20 and 40 generations) with q_t and Δq_t calculated every generation. Each simulation began with $q_o = 0.5$ and for each combination of population size, m and time period (NGENS), one hundred replicate simulations were made with different pseudo-random numbers in order that each mean estimate of b should have a reasonably small standard error. The results showed that -b appeared to increase linearly with m and decrease non-linearly with increasing time periods. However, population size had no detectable effect. Since the deviation between initial and equilibrium gene frequencies may also affect b, this possibility was checked because in real populations it is unlikely that q_o will be the same as $\bar{q}$. A second set of simulations was made with conditions the same as before except that only one population size (100 genes) was used and simulations started at a range of gene frequencies (0.1 to 0.9). Part of the results are given in Table 1 (as the results for $q_o > 0.5$ are a reflection of those for $q_o < 0.5$, they have been left out). It is clear that -b is a highly biased estimator of m even for long time periods and that the bias depends upon q_o, m and NGENS in a complex fashion. The bias is most severe for small NGENS and always declines as NGENS increases. However, the effect upon the bias of altering q_o depends upon m because the bias tends to decrease slightly as q_o increases when m=0, but tends to increase with

TABLE 2

Estimation of equilibrium gene frequency ($\hat{q}$) for various initial gene frequencies and immigration rates, when $\bar{q} = 0.5$.

q_o	m	NGENS		
		10	20	40
.1	0	.0927 ± .0131	.1462 ± .0322	.0757 ± .0130
	.1	.4239 ± .0557	.0281 ± .3643	.4858 ± .0087
	.2	.6274 ± .1143	.4893 ± .0067	.5057 ± .0044
.2	0	.4060 ± .1954	.0927 ± .0449	.1773 ± .0470
	.1	.3285 ± .0612	.4863 ± .0369	.6159 ± .0884
	.2	.4698 ± .0197	.5074 ± .0096	.4947 ± .0039
.3	0	.2487 ± .1026	.3395 ± .0486	.3094 ± .0438
	.1	.3829 ± .0247	.3824 ± .0684	.4955 ± .0097
	.2	.4862 ± .0170	.4900 ± .0163	.5001 ± .0039
.4	0	.2870 ± .0689	.3697 ± .0346	.3735 ± .0429
	.1	.4625 ± .0261	.5000 ± .0129	.4838 ± .0075
	.2	.4407 ± .0551	.4283 ± .0687	.4965 ± .0038
.5	0	.4641 ± .0735	.5405 ± .0559	.4824 ± .0358
	.1	.4991 ± .0156	.4701 ± .0275	.5026 ± .0085
	.2	.4968 ± .0077	.5264 ± .0122	.4990 ± .0041

q_o = initial gene frequency; m = immigration rate; NGENS = number of generations over which the regression is calculated. For each estimate, 100 replicate simulations were run and the standard error of each estimate is given.

q_o when m=0.2. This leads to the unexpected finding that when q_o and m are small, an increase in m may lead to a decrease in -b. When q_o is near 0.5 and especially for large NGENS, an increase in m will give an almost equivalent increase in -b. For some intermediate situations (e.g. q_o = 0.3, NGENS = 10) a change in m has no statistically significant effect on b. Further simulations of population with overlapping generations showed that if there is a lag between the time of mating and the time at which the progeny enter the observable population (this will occur, for example, in populations of animals where trapping misses the juveniles or plants with variable seed dormancies) then b is affected by the lag and the turnover of the population between censuses.

These simulation results show that many factors affect the bias in the estimation of m from b and in any realistic situation it will probably not be possible to compensate for this bias. This conclusion also applies to estimates of selection coefficients derived from b (Tamarin and Krebs, 1969); such estimates will be biased and no valid conclusions can be drawn from them.

Does the regression also yield biased estimates of the equilibrium gene frequency? Simulation results indicate that it does. In the second set of simulations described above, the estimated equilibrium gene frequency ($\hat{q}$) was obtained for each simulation from the equation,

$$\hat{q} = \bar{q}_t - \bar{\Delta q}_t/b \qquad \ldots\ldots (6)$$

where $\bar{q}_t$ and $\bar{\Delta q}_t$ are the mean values of q_t and Δq_t respectively. Some results are given in Table 2. When m = 0, $\hat{q}$ provided an apparently unbiased estimate of q_o, but when there is immigration, $\hat{q}$ appears to be an estimate of $\bar{q}$. However, for small m and NGENS the estimate $\hat{q}$ is biased towards q_o and hence this method will not yield an unbiased estimate of $\bar{q}$ unless m, or NGENS, or both are sufficiently large.

However, $\hat{q}$ may be a better estimate of $\bar{q}$ than an average of gene frequencies over all censuses. The two most likely averages to be considered are (i) the mean over all censuses, which will be very similar to $\bar{q}_t$ except that it will also include the last census, which is left out of the calculation of $\bar{q}_t$; (ii) if there are overlapping generations then many individuals will be censused more than once and it will be preferable to count each individual censused only once and calculate a mean frequency over all individuals. If $\hat{q}$ is biased towards q_o then from equation (6) (noting that usually b is negative and $\bar{\Delta q}_t$ is positive or negative as q_o is less than or greater than $\bar{q}$), $\bar{q}_t$ is biased towards q_o even more than $\hat{q}$. It is not easy to construct a simple argument on the bias of the two averages mentioned above, but by analogy with the bias of $\bar{q}_t$ it is likely that they are more biased than $\hat{q}$. If this assertion is true it implies that $\hat{q}$, biased though it may be, is the best estimate of $\bar{q}$ available until an improved estimator can be derived using a modification of this method or a new method.

To conclude, these results show that the regression of Δq on q should not be used to estimate immigration rates or selection pressures in populations and that the equilibrium gene frequencies estimated from the regression will be biased towards the initial gene frequency in the sequence of censuses.

Acknowledgements: Professor J.H. Bennett and Dr. O. Mayo made useful comments on the manuscript. I was supported by a University of Adelaide Research Grant Scholarship.

REFERENCES

KIMURA, M. and WEISS, G.H. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics, 49*, 561-576.

LI, C.C. 1955. The stability of an equilibrium and the average fitness of a population. *Amer. Natur., 89*, 281-295.

TAMARIN, R.H. and KREBS, C.J. 1969. *Microtus* population biology. II. Genetic changes at the transferrin locus in fluctuating populations of two vole species. *Evolution, 23*, 183-211.

TAMARIN, R.H. and KREBS, C.J. 1973. Selection at the transferrin locus in cropped vole populations. *Heredity, 30*, 53-62.

WILLIAMSON, M. 1972. *The Analysis of Biological Populations*. Edward Arnold, London.

APPENDIX 4

ANOTHER VIEW OF NEUTRAL ALLELES IN NATURAL POPULATIONS

Published in: *"Nature"* Volume 241; pp.463-464, 1973.

Kirby, G. C. & Halliday, R. B. (1973). Another view of neutral alleles in natural populations. *Nature, 241*(5390), 463-464.

BIBLIOGRAPHY

ANDERSON, P.K. (1964) Lethal alleles in *Mus musculus*: local distribution and evidence for isolation of demes. *Science*, 145: 177-178.

ANDERSON, P.K. (1970) Ecological structure and gene flow in small mammals. In *Variation in Mammalian Populations*. (Editors: R.J. Berry and H.N. Southern) Academic Press.

ANDERSON, P.K., L.C. DUNN and A.B. BEASLEY (1964) Introduction of a lethal allele into a feral house mouse population. *Amer. Nat.*, 98: 57-64.

ANDERSON, P.K. and J.L. HILL (1965) *Mus musculus*: Experimental induction of territory formation. *Science*, 148: 1753-1755.

ARNOLD, R. (1969) The effects of selection by climate on the land-snail *Cepaea nemoralis* (L.) *Evolution*, 23: 370-378.

ASHTON, G.C. and A.W. BRADEN (1961) Serum β-globulin polymorphism in mice. *Aust. J. Biol. Sci.*, 14: 248-253.

AYALA, F.J., J.R. POWELL and M.L. TRACEY (1972) Enzyme variability in the *Drosophila willistoni* group. V. Genic variation in natural populations of *Drosophila equinoxialis*. *Genet. Res.*, 20: 19-42

BADER, R.S. and W.H. LEHMANN (1965) Phenotypic and genotypic variation in odontometric traits of the house mouse. *Amer. Midl. Nat.*, 74: 28-38.

BAKER, R.H. (1946) A study of rodent populations on Guam, Mariana Islands. *Ecol. Monogr.*, 16: 393-408.

BARRAI, I. (1971) Subdivision and inbreeding. *Amer. J. Hum. Genet.*, 23: 95-96.

BERRY, R.J. (1963) Epigenetic polymorphism in wild populations of *Mus musculus*. *Genet. Res.*, 4: 193-220.

BERRY, R.J. (1964) The evolution of an island population of the house mouse. *Evolution*, 18: 468-483.

BERRY, R.J. (1968) The ecology of an island population of the house mouse. *J. Anim. Ecol.*, 37: 445-470.

BERRY, R.J. (1970) The natural history of the house mouse. *Fld. Stud.*, 3: 219-262.

BERRY, R.J. and H.M. MURPHY (1970) The biochemical genetics of an island population of the house mouse. *Proc. Roy. Soc. Lond.* B, 176: 87-103.

BIDDLE, F.G. and H. KRASNY (1970) The association of carbonic anhydrase activity with the non-hemoglobin erythrocytic protein (Pro-1) of the house mouse. *Can. J. Genet. Cytol.*, 12: 374.

BIDDLE, F.G. and M.L. PETRAS (1967) The inheritance of a non-hemoglobin erythrocytic protein in *Mus musculus*. *Genetics*, 57: 943-949.

BISHOP, J.A. (1972) An experimental study of the cline of industrial melanism in *Biston betularia* (L.) (Lepidoptera) between urban Liverpool and rural North Wales. *J. Anim. Ecol.*, 41: 209-243.

BODMER, W.F. and L.L. CAVALLI-SFORZA (1968) A migration matrix model for the study of random genetic drift. *Genetics*, 59: 565-592.

BREAKEY, D.R. (1963) The breeding season and age structure of feral house mouse populations near San Francisco Bay, California. *J. Mammal.*, 44: 153-168.

BROWN, A.H.D. (1970) The estimation of Wright's fixation index from genotypic frequencies. *Genetica*, 41: 399-406.

BROWN, L.N. (1965) Selection in a population of house mice containing mutant individuals. *J. Mammal.*, 46: 461-465.

BROWN, R.Z. (1953) Social behaviour, reproduction and population changes in the house mouse (*Mus musculus*, L.) *Ecol. Monogr.*, 23: 217-240.

BRUCK, D. (1957) Male segregation ratio advantage as a factor in maintaining lethal alleles in wild populations of house mice. *Proc. Nat. Acad. Sci.*, 43: 152-158.

BULMER, M.G. (1972) Multiple niche polymorphism. *Amer. Nat.*, 106: 254-257.

CAIN, A.J. and J.D. CURREY (1963) Area effects in *Cepaea*. *Phil. Trans. Roy. Soc. Lond.* B, 246: 1-81.

CAIN, A.J. and P.M. SHEPPARD (1954) Natural selection in *Cepaea*. *Genetics*, 39: 89-116.

CANNINGS, C. and A.W.F. EDWARDS (1969) Expected genotypic frequencies in a small sample: deviations from Hardy-Weinberg equilibrium. *Amer. J. Hum. Genet.*, 21: 245-247.

CAVALLI-SFORZA, L.L., I. BARRAI and A.W.F. EDWARDS (1964) Analysis of human evolution under random genetic drift. *Cold Spring Harbor Symp. Quant. Biol.*, 29: 9-20.

CHRISTIAN, J.J. (1963) The pathology of overpopulation. *Mil. Med.*, 128: 571-603.

CHRISTIANSEN, F.B. (1974) Sufficient conditions for protected polymorphism in a subdivided population. *Amer. Nat.*, 108: 157-166.

CLARKE, B. (1966) The evolution of morph-ratio clines. *Amer. Nat.*, 100: 389-402.

CLARKE, B. (1970) Darwinian evolution of proteins. *Science*, 168: 1009-1011.

COCKERHAM, C.C. (1967) Group inbreeding and coancestry. *Genetics*, 56: 89-104.

COCKERHAM, C.C. (1969) Variance of gene frequencies. *Evolution*, 23: 72-84.

COCKERHAM, C.C. (1973) Analyses of gene frequencies. *Genetics*, 74: 679-700.

COHEN, B.L. and D.C. SHREFFLER (1961) A revised nomenclature for the mouse transferrin locus. *Genet. Res.*, 2: 306-308.

COOK, L.M. (1971) "Coefficients of natural selection." Hutchinson: London.

CROSBY, J.L. (1960) The use of electronic computation in the study of random fluctuations in rapidly evolving populations. *Phil. Trans. Roy. Soc.*, 242: 551-573.

CROW, J.F. (1970) Genetic loads and the cost of natural selection. In *Mathematical Topics in Population Genetics* (Editor: K. Kojima) pp.128-177. Springer-Verlag: Berlin.

CROW, J.F. and M. KIMURA (1970) *An Introduction to Population Genetics Theory*. Harper and Row: New York.

CROWCROFT, P. (1955) Territoriality in wild house mice, *Mus musculus*, L. *J. Mammal.*, 36: 299-301.

CROWCROFT, P. (1966) *Mice All Over*. G.T. Foulis & Co. Ltd.: London.

CROWCROFT, P. and F.P. ROWE (1957) The growth of confined colonies of the wild house mouse (*Mus musculus*, L.). *Proc. Zool. Soc. Lond.*, 129: 359-370.

CROWCROFT, P. and F.P. ROWE (1958) The growth of confined colonies of wild house mouse: the effect of dispersal on female fecundity. *Proc. Zool. Soc. Lond.*, 131: 357-365.

CROWCROFT, P. and F.P. ROWE (1963) Social organisation and territorial behaviour in the wild house mouse (*Mus musculus*, L.). *Proc. Zool. Soc. Lond.*, 140: 517-531.

DEAKIN, M.A.B. (1966) Sufficient conditions for genetic polymorphism. *Amer. Nat.*, 100: 690-691.

DEAKIN, M.A.B. (1968) Genetic polymorphism in a subdivided population. *Aust. J. Biol. Sci.*, 21: 165-168.

DEAKIN, M.A.B. (1972) Corrigendum to genetic polymorphism in a subdivided population. *Aust. J. Biol. Sci.*, 25: 213-214.

De FRIES, J.C. and G.E. McCLEARN (1970) Social dominance and Darwinian fitness in the laboratory mouse. *Amer. Nat.*, 104: 408-411.

De FRIES, J.C. and G.E. McCLEARN (1972) Behavioural genetics and the fine structure of mouse populations: a study in microevolution. In *Evolutionary Biology*, 5: 279-291.

De LONG, K.T. (1967) Population ecology of feral house mice. *Ecology*, 48: 611-634.

DICKINSON, H. and J. ANTONOVICS (1973) Theoretical considerations of sympatric divergence. *Amer. Nat.*, 107: 256-274.

DOBZHANSKY, Th. (1973) Active dispersal and passive transport in Drosophila. *Evolution*, 27: 565-575.

DOBZHANSKY, Th. and S. WRIGHT (1941) Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura*. *Genetics*, 26: 23-51.

DOBZHANSKY, Th. and S. WRIGHT (1943) Genetics of natural populations. X. Dispersion rates in *Drosophila pseudoobscura*. *Genetics*, 28: 304-340.

DUNN, L.C., A.B. BEASLEY and H. TINKER (1958) Relative fitness of wild house mice heterozygous for a lethal allele. *Amer. Nat.*, 92: 215-220.

DUNN, L.C., A.B. BEASLEY and H. TINKER (1960) Polymorphisms in populations of wild house mice. *J. Mammal.*, 41: 220-229.

DUNN, L.C. and D. BENNETT (1966) Report on viability. *Mouse News Letter*, 34: 20-22.

DUNN, L.C. and H. LEVENE (1961) Population dynamics of a variant *t* allele in a confined population of wild house mice. *Evolution*, 15: 385-393.

EDWARDS, A.W.F. (1971) Distances between populations on the basis of gene frequencies. *Biometrics*, 27: 873-881.

EVANS, F.C. (1949) A population study of house mice (*Mus musculus*) following a period of local abundance. *J. Mammal.*, 30: 351-363.

EWENS, W.J. (1969) *Population Genetics* Methuen: London.

FELSENSTEIN, J. (1971) Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics*, 68: 581-597.

FISHER, R.A. (1922) On the dominance ratio. *Proc. Roy. Soc. Edinburgh*, 42: 321-341.

FISHER, R.A. (1930) *The Genetical Theory of Natural Selection*. Clarendon Press: Oxford.

FISHER, R.A. (1937) The wave of advance of advantageous genes. *Ann. Eugen.*, 7: 355-369.

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

FISHER, R.A. (1958) *The Genetical Theory of Natural Selection* 2nd edition. Dover Press: New York.

FISHER, R.A. and E.B. FORD (1947) The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity*, 1: 143-174.

FORD, E.B. (1964) *Ecological Genetics* Methuen: London.

FRANKLIN, I. and R.C. LEWONTIN (1970) Is the gene the unit of selection? *Genetics* 65: 707-734.

FRASER, A. and D. BURNELL (1970) *Computer Models in Genetics* McGraw-Hill, Inc.

GAHNE, B. (1966) Studies on the inheritance of electrophoretic forms of transferrin, albumins, prealbumins and plasma esterases of horses. *Genetics*, 53: 681-694.

GAINES, M.S. and C.J. KREBS (1971) Genetic changes in fluctuating vole populations. *Evolution*, 25: 702-723

GIBLETT, E.R. (1969) *Genetic Markers in Human Blood* Blackwell: Oxford.

GILLESPIE, J. (1974) Polymorphism in patchy environments. *Amer. Nat.*, 108: 145-151.

GOULD, S.J. and R.F. JOHNSTON (1972) Geographic variation. *Ann. Rev. Ecol. System.*, 3: 457-498.

GREEN, M.C. (1966) Mutant genes and linkages. In *Biology of the Laboratory Mouse* 2nd edition. (Ed.: E.L. Green) McGraw-Hill: New York.

HALDANE, J.B.S. (1927) A mathematical theory of natural and artificial selection. V. Selection and mutation. *Proc. Camb. Phil. Soc.*, 28: 838-844.

HALDANE, J.B.S. (1930) A mathematical theory of natural and artificial selection. VI. Isolation. *Proc. Camb. Phil. Soc.*, 26: 220-230.

HALDANE, J.B.S. (1932) *The Causes of Evolution* Harper and Row: New York.

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

HALDANE, J.B.S. (1954) An exact test for randomness of mating. *J. Genet.*, 52: 631-635.

HANSON, 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.

HARRIS, H., D.A. HOPKINSON and E.R. ROBSON (1974) The incidence of rare alleles determining electrophoretic variants: data on 43 enzyme loci in man. *Ann. Hum. Genet.*, 37: 237-253.

HEDRICK, P.W. (1971) A new approach to measuring genetic similarity. *Evolution*, 25: 276-280.

HOPKINSON, D.A., M.A. MESTRINER, J. CORTNER and H. HARRIS (1973) Esterase D: a new human polymorphism. *Ann. Hum. Genet. Lond.*, 37: 119-137.

HUNT, W.G. and R.K. SELANDER (1973) Biochemical genetics of hybridisation in European house mice. *Heredity*, 31: 11-33.

HUTTON, J.J. (1969) Linkage analyses using biochemical variants in mice. I. Linkage of the hemoglobin beta-chain and glucosephosphate isomerase loci. *Biochem. Genet.*, 3: 507-515.

HUTTON, J.J., J. BISHOP, R. SCHECT and E.S. RUSSELL (1962) Hemoglobin inheritance in inbred mouse strains. I. Structural differences. *Proc. Nat. Acad. Sci.*, 48: 1505-1513.

HUXLEY, J.S. (1942) *Evolution. The Modern Synthesis.* Allen and Unwin: London.

IMAIZUMI, Y., N.E. MORTON and D.E. HARRIS (1970) Isolation by distance in artificial populations. *Genetics*, 66: 569-582.

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

JAIN, S.K. and P.L. WORKMAN (1967) Generalised F-statistics and the theory of inbreeding and selection. *Nature*, 214: 674-678.

JAMES, S.H. (1970) A demonstration of a possible mechanism of sympatric divergence using simulation techniques. *Heredity*, 25: 241-252.

JOHNSON, G.B. (1972) Evidence that enzyme polymorphisms are not selectively neutral. *Nature New Biology*, 237: 170-171.

JOHNSON, M.S. (1971) Adaptive lactate dehydrogenase variation in the crested blenny, *Anoplarchus*. *Heredity*, 27: 205-226.

JOHNSTON, P.G. and G.H. BROWN (1969) A comparison of the relative fitness of genotypes segregating for the t^{w2} allele in laboratory stock and its possible effect on gene frequency in mouse populations. *Amer. Nat.*, 103: 5-21.

JOLLY, G.M. (1965) Explicit maximum likelihood estimates from capture-recapture data with both death and immigration - stochastic model. *Biometrika*, 52: 225-247.

KARLIN, S. and J. McGREGOR (1972) Application of method of small parameters to multi-niche population genetic models. *Theor. Pop. Biol.*, 3: 186-209.

KETTLEWELL, H.B.D. and R.J. BERRY (1961) The study of a cline. *Amathes glareosa* Esp. and its melanic *f. edda* Staud (Lep.) in Shetland. *Heredity*, 16: 403-414.

KIMURA, M. (1953) "Stepping stone" model of population. *Ann. Rept. Nat. Inst. Genetics, Japan*, 3: 62-63.

KIMURA, M. (1964) Diffusion models in population genetics. *J. Appl. Prob.*, 1: 177-232.

KIMURA, M. (1968) Evolutionary rate at the molecular level. *Nature*, 217: 624-626.

KIMURA, M. (1970) Stochastic processes in population genetics, with special reference to distribution of gene frequencies and probability of gene fixation. In *Mathematical Topics in Population Genetics* (Editor: K. Kojima). Springer-Verlag: Berlin.

KIMURA, M. and J.F. CROW (1963) The measurement of effective population number. *Evolution*, 17: 279-288.

KIMURA, M. and J.F. CROW (1964) The number of alleles that can be maintained in a finite population. *Genetics*, 49: 725-738.

KIMURA, M. and T. MARUYAMA (1971) Pattern of neutral polymorphism in a geographically structured population. *Genet. Res.*, 18: 125-131.

KIMURA, M. and T. OHTA (1971a) Protein polymorphism as a phase of molecular evolution. *Nature*, 229: 467-469.

KIMURA, M. and T. OHTA (1971b) *Theoretical Aspects of Population Genetics*. Princeton University Press: Princeton.

KIMURA, M. and G.H. WEISS (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49: 561-576.

KING, J.L. and T.H. JUKES (1969) Non-Darwinian evolution: random fixation of selectively neutral mutations. *Science*, 164: 788-798.

KIRBY, G.C. (1973) Heterozygote frequencies in small house-mouse populations. Abst. *Proc. XIIIth Int. Cong. Genet.*, Berkeley. *Genetics* (Suppl.), 74: S140-S141.

KIRBY, G.C. and R.B. HALLIDAY (1973) Another view of neutral alleles in natural populations. *Nature*, 241: 463-464.

KOEHN, R.K. (1969) Esterase heterogeneity: dynamics of a polymorphism. *Science*, 163: 943-944.

KOJIMA, K. and K.M. YARBROUGH (1967) Frequency- dependent selection at the esterase-6 locus in *Drosophila melanogaster*. *Proc. Nat. Acad. Sci. U.S.*, 57: 645-649.

LAMOTTE, M. (1959) Polymorphism of natural populations of *Cepaea nemoralis*. *Cold Spring Harbor Symp. Quant. Biol.*, 24: 65-86.

LATTER, B.D.H. (1973a) The island model of population differentiation: a general solution. *Genetics*, 73: 147-157.

LATTER, B.D.H. (1973b) The estimation of genetic divergence between populations based on gene frequency data. *Amer. J. Hum. Genet.*, 25: 247-261.

LAURIE, E.M.O. (1946) The reproduction of the house-mouse (*Mus musculus*) living in different environments. *Proc. Roy. Soc.*, B, 133: 268-281.

LERNER, I.M. (1954) *Genetic Homeostasis*. Oliver and Boyd: Edinburgh.

LEVENE, H. (1949) On a matching problem arising in genetics. *Ann. Math. Statist.*, 20: 91-94.

LEVENE, H. (1953) Genetic equilibrium when more than one ecological niche is available. *Amer. Nat.*, 87: 331-333.

LEVIN, B.R., M.L. PETRAS and D.I. RASMUSSEN (1969) The effect of migration on the maintenance of a lethal polymorphism in the house mouse. *Amer. Nat.*, 103: 647-661.

LEVINS, R. (1968) *Evolution in Changing Environments*. Princeton University Press: Princeton.

LEWONTIN, R.C. (1962) Interdeme selection controlling a polymorphism in the house mouse. *Amer. Nat.*, 96: 65-78.

LEWONTIN, R.C. (1967) Population genetics. *Annual Review of Genetics*, 1: 37-70.

LEWONTIN, R.C. (1968) The effect of differential viability on the population dynamics of *t* alleles in the house mouse. *Evolution*, 22: 262-273.

LEWONTIN, R.C. and L.C. DUNN (1960) The evolutionary dynamics of a polymorphism in the house mouse. *Genetics*, 45: 705-722.

LEWONTIN, R.C. and J.L. HUBBY (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 54: 595-609.

LEWONTIN, R.C. and J. KRAKAUER (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 74: 175-195.

LI, C.C. (1955) *Population Genetics* University of Chicago Press: Chicago.

LI, C.C. (1969) Population subdivision with respect to multiple alleles. *Ann. Hum. Genet.*, 33: 23-29.

LIDICKER, W.Z. (1962) Emigration as a possible mechanism permitting the regulation of population density below carrying capacity. *Amer. Nat.*, 96: 29-34.

LIDICKER, W.Z. (1966) Ecological observations on a feral house mouse population declining to extinction. *Ecol. Monogr.*, 36: 27-50.

LINCOLN, F.C. (1930) Calculating waterfowl abundance on the basis of banding returns. *U.S.D.A. Circu.*, 118: 1-4.

LLOYD, J.A. and J.J. CHRISTIAN (1967) Relationship of activity and agression to density in two confined populations of house-mice (*Mus musculus*). *J. Mammal.*, 48: 262-269.

LLOYD, J.A. and J.J. CHRISTIAN (1969) Reproductive activity of individual females in three experimental freely growing populations of house mice (*Mus musculus*). *J. Mammal.*, 50: 49-59.

MALECOT, G. (1965) Identical loci and relationship. *Vth Berkeley Symposium on Probability and Statistics*, 317-332.

MALECOT, G. (1969) *The Mathematics of Heredity* Translated by D.M. Yermanos. W.H. Freeman & Co.: San Francisco.

MANWELL, C. and C.M.A. BAKER (1970) *Molecular Biology and the Origin of Species.* Sidgwick and Jackson: London.

MARTIN, J.E. and M.L. PETRAS (1971) Es-3 esterases in erythrocytes of *Mus musculus*. *Can. J. Genet. Cytol.*, 13: 777-781.

MARUYAMA, T. (1969) Genetic correlation in the stepping stone model with non-symmetrical migration rates. *J. Appl. Prob.*, 6: 463-477.

MARUYAMA, T. (1970a) Rate of decrease of genetic variability in a subdivided population. *Biometrika*, 57: 299-311.

MARUYAMA, T. (1970c) On the rate of decrease of heterozygosity in circular stepping stone models of populations. *Theor. Pop. Biol.*, 1: 101-119.

MARUYAMA, T. (1970d) Stepping stone models of finite length. *Adv. Appl. Prob.*, 2: 229-258.

MARUYAMA, T. (1970e) Analysis of population structure. I. One-dimensional stepping stone models of finite length. *Ann. Hum. Genet. Lond.*, 34: 201-219.

MARUYAMA, T. (1970f) Effective number of alleles in a subdivided population. *Theor. Pop. Biol.*, 1: 273-306.

MARUYAMA, T. (1971) Analysis of population structure. II. Two-dimensional stepping stone models of finite length and other geographically structured populations. *Ann. Hum. Genet. Lond.*, 35: 179-196.

MARUYAMA, T. (1972) Distribution of gene frequencies in a geographically structured finite population. I. Distribution of neutral genes and of genes with small effect. *Ann. Hum. Genet.*, 35: 411-423.

MATHER, K. (1973) *Genetical Structure of Populations*. Chapman and Hall: London.

MAYNARD SMITH, J. (1966) Sympatric speciation. *Amer. Nat.*, 100: 637-650.

MAYNARD SMITH, J. (1968) *Mathematical Ideas in Biology*. Cambridge University Press: Cambridge.

MAYNARD SMITH, J. (1970) Genetic polymorphism in a varied environment. *Amer. Nat.*, 104: 487-490.

MAYR, E. (1963) *Animal Species and Evolution*. Harvard University Press: Cambridge.

MILKMAN, R. (1973) Electrophoretic variation of *Escherichia coli* from natural sources. *Science*, 182: 1024-1026.

MORAN, P.A.P. (1959) The theory of some genetical effects of population subdivision. *Aust. J. Biol. Sci.*, 12: 109-116.

MORAN, P.A.P. (1962) *The Statistical Processes of Evolutionary Theory*. Clarendon Press: Oxford.

MORTON, J.R. (1962) Starch-gel electrophoresis of mouse haemoglobins. *Nature*, 194: 383-384.

MORTON, N.E. (1969a) Population structure. In *Computer Applications in Genetics* (Editor: N.E. Morton), pp.61-68. University of Hawaii Press.

MORTON, N.E. (1969b) Human population structure. *Annual Review of Genetics*, 3: 53-74.

NASSAR, R., H.J. MUHS and R.D. COOK (1973) Frequency-dependent selection at the Payne inversion in *Drosophila melanogaster*. *Evolution*, 27: 558-564.

NEEL, J.V. and R.H. WARD (1972) The genetic structures of a tribal population, the Yanomama Indians. VI. Analysis by F-statistics (including a comparison with the Makiritare and Xavante). *Genetics*, 72: 639-666.

NEI, M. (1965) Variation and covariation of gene frequencies in subdivided populations. *Evolution*, 19: 256-258.

NEI, M. (1972) Genetic distance between populations. *Amer. Nat.*, 6: 283-292.

NEI, M. (1973) Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci.*, 70: 3321-3323.

NEI, M. and Y. IMAIZUMI (1966) Genetic structure of human populations. I. Local differentiation of blood group gene frequencies in Japan. *Heredity*, 21: 9-35.

NEWSOME, A.E. (1969a) A population study of house-mice temporarily inhabiting a South Australian wheatfield. *J. Anim. Ecol.*, 38: 341-359.

NEWSOME, A.E. (1969b) A population study of house-mice permanently inhabiting a reed-bed in South Australia. *J. Anim. Ecol.*, 38: 361-377.

NEWSOME, A.E. (1970) An experimental attempt to produce a mouse plague. *J. Anim. Ecol.*, 39: 299-311.

NEWSOME, A.E. (1971) The ecology of house-mice in cereal haystacks. *J. Anim. Ecol.*, 40: 1-15.

NEWSOME, A.E. and P. CROWCROFT (1971) Outbreaks of house-mice in South Australia in 1965. *CSIRO Wildlife Research*, 16: 41-47.

OAKESHOTT, J.G. (1974) Social dominance, aggressiveness and mating success among male house mice (*Mus musculus*). *Oecologia*, 15: 143-158.

O'GOWER, A.K. and P.I. NICOL (1968) A latitudinal cline of haemoglobins in a bivalve mollusc. *Heredity*, 23: 485-491.

OWEN, J.A. and H. SMITH (1961) Detection of ceruloplasmin after zone electrophoresis. *Clin. Chim. Acta*, 6: 441-444.

PARSONS, P.A. (1963) Migration as a factor in natural selection. *Genetica*, 33: 184-206.

PEARSON, O.P. (1963) History of two local outbreaks of feral house mice. *Ecology*, 44: 540-549.

PELZER, C.F. (1965) Genetic control of erythrocytic esterase forms in *Mus musculus*. *Genetics*, 52: 819-828.

PETRAS, M.L. (1963) Genetic control of a serum esterase component in *Mus musculus*. *Proc. Nat. Acad. Sci.*, 50: 112-116.

PETRAS, M.L. (1967a) Studies of natural populations of *Mus*. I. Biochemical polymorphisms and their bearing on breeding structure. *Evolution*, 21: 259-274.

PETRAS, M.L. (1967b) Studies of natural populations of *Mus*. II. Polymorphism at the *T* locus. *Evolution*, 21: 466-478.

PETRAS, M.L. (1967c) Studies of natural populations of *Mus*. III. Coat colour polymorphisms. *Can. J. Genet. Cytol.*, 9: 287-296.

PETRAS, M.L. and F.G. BIDDLE (1967) Serum esterases in the house mouse, *Mus musculus*. *Can. J. Genet. Cytol.*, 9: 704-710.

PETRAS, M.L. and J.E. MARTIN (1969) Improved electrophoretic resolution of some hemoglobin variants in *Mus musculus*. *Biochem. Genet.*, 3: 303-309.

PETRAS, M.L., J.D. REIMER, F.G. BIDDLE, J.E. MARTIN and R.S. LINTON (1969) Studies of natural populations of *Mus*. V. A survey of nine loci for polymorphisms. *Can. J. Genet. Cytol.*, 11: 497-513.

PETRAS, M.L. and P. SINCLAIR (1969) Another esterase variant in the kidney of the House mouse, *Mus musculus*. *Can. J. Genet. Cytol.*, 11: 97-102.

PIHAR, O. (1968) Localization of carbonate hydro-lyase activity in starch gel electrophoresis. *Z. Klin. Chem. u. Klin. Biochem.*, 2: 110-111.

POLLAK, E. (1968) On random drift in a subdivided population. *J. Appl. Prob.*, 5: 314-333.

POPP, R.A. (1962) Studies on the mouse haemoglobin loci. V. Differences among tryptic peptides of the β-chain governed by alleles at the Hb locus. *J. Hered.*, 53: 142-146.

POPP, R.A. (1966) Inheritance of an erythrocytic and kidney esterase in the mouse. *J. Hered.*, 57: 197-201.

POPP, R.A. and D.M. POPP (1962) The inheritance of serum esterases having different electrophoretic patterns among inbred lines of mice. *J. Hered.*, 53: 111-114.

POPP, R.A. and W. St. AMAND (1960) Studies on the mouse hemoglobin locus. I. Identification of hemoglobin types and linkage of hemoglobin with albinism. *J. Hered.*, 51: 141-144.

PROUT, T. (1968) Sufficient conditions for multiple niche polymorphism. *Amer. Nat.*, 102: 493-496.

QUADAYNO, D.M. (1968) Home range size in feral house mice. *J. Mammal.*, 49: 149-151.

RANNEY, H.M. and S. GLUECKSON-WAELSCH (1955) Filter-paper electrophoresis of mouse haemoglobins: preliminary note. *Ann. Human Genet.*, 19: 269-272.

RASMUSSEN, D.I. (1964) Blood group polymorphism and inbreeding in natural populations of the deer mouse *Peromyscus maniculatus gracilis*. *Evolution*, 18: 219-229.

REIMER, J.D. and M.L. PETRAS (1967) Breeding structure of the house mouse, *Mus musculus*, in a population cage. *J. Mammal.*, 48: 88-99.

ROBERTSON, A. (1965) The interpretation of genotypic ratios in domestic animal populations. *Animal Production*, 7: 319-324.

ROBINSON, R. (1971) *Gene Mapping in Laboratory Mammals* Part A. Plenum Press: London.

ROHLF, F.J. and G.D. SCHNELL (1971) An investigation of the isolation-by-distance model. *Amer. Nat.*, 105: 295-324.

ROWE, F.P., E.J. TAYLOR and A.H.J. CHUDLEY (1963) The numbers and movements of house-mice (*Mus musculus*, L.) in the vicinity of four corn-ricks. *J. Anim. Ecol.*, 32: 87-97.

ROWE, F.P., E.J. TAYLOR and A.H.J. CHUDLEY (1964) The effect of crowding on the reproduction of the house-mouse (*Mus musculus*) living in corn-ricks. *J. Anim. Ecol.*, 33: 477-483.

RUDDLE, F.H. and T.H. RODERICK (1965) The genetic control of three kidney esterases in C57BL/6J and RF/J mice. *Genetics*, 51: 445-454.

RUDDLE, F.H., T.H. RODERICK, T.B. SHOWS, P.G. WEIGL, R.K. CHIPMAN and P.K. ANDERSON (1969) Measurement of genetic heterogeneity by means of enzyme polymorphisms in wild populations of the mouse. *J. Hered.*, 60: 321-322.

RUDDLE, F.H., T.B. SHOWS and T.H. RODERICK (1969) Esterase genetics in *Mus musculus*: expression, linkage and polymorphism of locus Es-2. *Genetics*, 62: 393-399.

SHREFFLER, D.C. (1963) Linkage of the mouse transferrin locus. *J. Hered.*, 54: 127-129.

SHREFFLER, D.C. (1964) Inheritance of a serum pre-albumin variant in the mouse. *Genetics*, 49: 629-634.

SCHOPF, T.J.M. and J.L. GOOCH (1971) Gene frequencies in a marine ectoproct: a cline in natural populations related to sea temperature. *Evolution*, 25: 286-289.

SCHWARZ, E. and H.K. SCHWARZ (1943) The wild and commensal stocks of the house mouse, *Mus musculus* Linnaeus. *J. Mammal.*, 24: 59-72.

SEARLE, A.G. (1974) Editorial. *Mouse News Letter*, 50: 1.

SEBER, G.A.F. (1973) *The Estimation of Animal Abundance and Related Parameters*. Charles Griffin: London.

SELANDER, R.K. (1970) Behaviour and genetic variation in natural populations. *Amer. Zool.*, 10: 53-66.

SELANDER, R.K., W.G. HUNT and S.Y. YANG (1969) Protein polymorphism and genic heterozygosity in two European subspecies of the house mouse. *Evolution*, 23: 379-390.

SELANDER, R.K. and D.W. KAUFMAN (1973) Self-fertilization and genetic population structure in a colonizing land snail. *Proc. Nat. Acad. Sci. U.S.*, 70: 1186-1190.

SELANDER, R.K. and S.Y. YANG (1969) Protein polymorphism and genic heterozygosity in a wild population of the house mouse. *Genetics*, 63: 653-667.

SELANDER, R.K. and S.Y. YANG (1970) Biochemical genetics and behaviour in wild house mouse populations. In *Contributions to Behaviour-Genetic Analysis, the Mouse as a Prototype*. (Editors: L. Gardner and D.D. Thiessen) Appleton-Century-Crofts: New York.

SELANDER, R.K., S.Y. YANG and W.G. HUNT (1969) Polymorphism in esterases and haemoglobin in wild populations of the House Mouse (*Mus musculus*). *Studies in Genetics*, V: 271-338.

SHAW, C.R. and R. PRASAD (1970) Starch gel electrophoresis of enzymes - a compilation of recipes. *Biochem. Genet.*, 4: 297-320.

SING, C.F., G.J. BREWER and B. THIRTLE (1973) Inherited biochemical variation in *Drosophila melanogaster*: noise or signal? I. Single locus analyses. *Genetics*, 75: 381-404.

SKELLAM, J.G. (1951) Random dispersal in theoretical populations. *Biometrika*, 38: 196-218.

SLATKIN, M. (1972) On treating the chromosome as the unit of selection. *Genetics*, 72: 157-168.

SMITH, C.A.B. (1969) Local fluctuations in gene frequencies. *Ann. Hum. Genet.*, 32: 251-260.

SMITH, C.A.B. (1970) A note on testing the Hardy-Weinberg law. *Ann. Hum. Genet.*, 33: 377-383.

SMITH, W.W. (1954) Reproduction in the house mouse, *Mus musculus* L., in Mississippi. *J. Mammal.*, 35: 509-515.

SMITHIES, O. (1959a) Zone electrophoresis in starch gels and its application to studies of serum proteins. In *Advances in Protein Chemistry* (Editors: Anfinson, C.B., Anson, M.L., Bailey, K. and J.T. Edsall), 14: 65. Academic Press: New York.

SMITHIES, O. (1959b) An improved procedure for starch gel electrophoresis: further variations in the serum proteins of normal individuals. *Biochem. J.*, 71: 585.

SNEDECOR, G.W. (1966) *Statistical Methods*, 5th edition. Iowa State University Press: Ames, Iowa.

SOUTHERN, H.N. (Editor) (1954) *Control of Rats and Mice. Vol. III House Mice.* Oxford University Press: London.

SOUTHERN, H.N. and E.M.O. LAURIE (1946) The house-mouse (*Mus musculus*) in corn ricks. *J. Anim. Ecol.*, 15: 134-149.

SOUTHWICK, C.H. (1955a) The population dynamics of confined house mice supplied with unlimited food. *Ecology*, 36: 212-225.

SOUTHWICK, C.H. (1955b) Regulatory mechanisms of house mouse populations: social behaviour affecting litter survival. *Ecology*, 36: 627-634.

SOUTHWICK, C.H. (1958) Population characteristics of house mice living in English corn ricks: density relationships. *Proc. Zool. Soc. Lond.*, 131: 163-175.

STRECKER, R.C. and J.T. EMLEN (1953) Regulatory mechanisms in house-mouse populations: the effect of limited food supply on a confined population. *Ecology*, 34: 375-385.

STROBECK, C. (1974) Sufficient conditions for polymorphism with N niches and M mating groups. *Amer. Nat.*, 108: 152-156.

TAMARIN, R.H. and S.R. MALECHA (1971) The population biology of Hawaiian rodents: demographic parameters. *Ecology*, 52: 383-394.

TEMPLETON, M. (1969) Mouse erythrocytic esterases and an interaction with heparin as shown by starch gel electrophoresis. *J. Histochem. Cytochem.*, 17: 95-101.

VAN VALEN, L. (1965) Selection in natural populations. IV. British house mice (*Mus musculus*). *Genetica*, 36: 119-134.

WEBER, W. (1950) Genetical studies on the skeleton of the mouse. III. Skeletal variation in wild populations. *J. Genet.*, 50: 174-178.

WHEELER, L.L. and R.K. SELANDER (1972) Genetic variation in populations of the House mouse, *Mus musculus*, in the Hawaiian islands. *Studies in Genetics*, VII: 269-296.

WHITAKER, J.O. (1966) Food of *Mus musculus*, *Peromyscus maniculatus bairdii* and *Peromyscus leucopus* in Vigo County, Indiana. *J. Mammal.*, 47: 473-486.

WRIGHT, S. (1921) Systems of mating. *Genetics*, 6: 111-178.

WRIGHT, S. (1922) Coefficients of inbreeding and relationship. *Amer. Nat.*, 56: 330-338.

WRIGHT, S. (1929) The evolution of dominance. Comment on Dr. Fisher's reply. *Amer. Nat.*, 63: 274-279.

WRIGHT, S. (1931) Evolution in Mendelian populations. *Genetics*, 16: 97-159.

WRIGHT, S. (1938) Size of populations and breeding structure in relation to evolution. *Science*, 87: 430-431.

WRIGHT, S. (1943a) Isolation by distance. *Genetics*, 28: 114-138.

WRIGHT, S. (1943b) An analysis of local variability of flower color in *Linanthus parryae*. *Genetics*, 28: 139-156.

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

WRIGHT, S. (1951) The genetical structure of populations. *Ann. Eugen.*, 15: 323-354.

WRIGHT, S. (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 19: 395-420.

WRIGHT, S. (1969) *Evolution and the Genetics of Populations*. Vol. 2 *The Theory of Gene Frequencies*. University of Chicago Press.

WRIGHT, S. (1970) Random drift and the shifting balance theory of evolution. *Biomathematics* Vol. 1 *Mathematical Topics in Population Genetics*. (Editor: K. Kojima) Springer-Verlag: Berlin.

WRIGHT, S. and Th. DOBZHANSKY (1946) Genetics of natural populations. XII Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. *Genetics*, 31: 125-256.

YAMAZAKI, T. and T. MARUYAMA (1973) Evidence that enzyme polymorphisms are selectively neutral. *Nature*, 245: 140-141.