

TERTIARY LEVEL BIOLOGY

Population Genetics

J. S. GALE

TERTIARY LEVEL BIOLOGY

Population Genetics

J. S. GALE

Lecturer in Genetics
University of Birmingham
England

Blackie

Glasgow and London

Blackie & Son Limited
Bishopbriggs
Glasgow G64 2NZ

Furnival House
14-18 High Holborn
London WC1V 6BX

© 1980 J. S. Gale
First published 1980

*All rights reserved.
No part of this publication may be reproduced,
stored in a retrieval system or transmitted,
in any form or by any means,
electronic, mechanical, recording or otherwise,
without prior permission of the Publishers*

*International Standard Book Numbers
Hardback 0 216 91000 5
Paperback 0 216 91001 3*

Filmset by Advanced Filmsetters (Glasgow) Limited
Printed in Great Britain by
Thomson Litho Ltd., East Kilbride, Scotland

Preface

POPULATION GENETICS IS OFTEN THOUGHT TO BE A DIFFICULT SUBJECT. To some extent, difficulties are inevitable in a field where some quite basic points are controversial. However, problems are most acute when theoretical points are discussed, despite the fact that there has been very little controversy over the mathematics. In my experience, the actual mathematical manipulations rarely cause much difficulty. Rather it is that the biologist, lacking the physicist's or chemist's experience in "reading" mathematical formulae, finds it difficult to appreciate what is happening in a mathematical treatment and to grasp the implications of the results obtained, when these are given in mathematical form. Accordingly, I have followed a procedure, which students seem to find helpful, of giving a rough-and-ready verbal treatment of a problem before attempting a much more exact mathematical treatment; when the results of the latter are not readily interpretable, I have given an elucidation. Another problem which often concerns students is the reliability of results obtained using approximate methods; I have, therefore, discussed this in fair detail in critical cases.

When dealing with controversial issues, I have done my very best to be fair. To conceal one's opinions entirely would probably make for a very dull book. I trust, however, that I have given enough for the reader previously unfamiliar with these controversies to form his own judgement.

To acknowledge all those who have so greatly assisted my understanding of population genetics would mean a very lengthy list. I should, however, particularly mention E. J. Machin, who introduced me to the subject in my schooldays, and the inspiring lectures of Dr. A. R. G. Owen. To express my debt to Sir Ronald Fisher would require literary powers far beyond my own; in his presence, indeed, "meadow, grove, and stream,/The earth, and every common sight/To me did seem/Apparelled in celestial light".

No author could be blessed with more helpful colleagues. I should particularly like to thank Dr. P. D. S. Caligari for his willingness to give unlimited time to discussing problems and their presentation, and for his many valuable suggestions. I am also indebted for encouragement, assistance and advice to Prof. J. L. Jinks, Rev. Dr. L. J. Eaves, Dr. M. J. Kearsey, Dr. A. J. Birley, Dr. G. H. Jones, Dr. A. J. Cornish-Bowden, Dr. N. Goodchild, Mr. I. J. Mackay and Mr. J. P. Gibson. Prof. Bryan Clarke was kind enough to read and comment on the section on *Cepaea* (although he would not necessarily agree with all that appears there). Finally, I should like to record my debt to my students. By their comments, queries, criticisms and (very occasional!) errors, they have guided me to what I believe to be a greatly improved understanding of the subject and its presentation. Of course, I take sole responsibility for anything in the book which is incompetent, irrelevant or immaterial.

J.S.G.

18	APPENDIXES
139	Chapter 10. POLYMERIZATION
150	Chapter 11. EVOLUTION OF ALTRUISM
161	Chapter 12. SUMMARY, SOME TENTATIVE CONCLUSIONS
164	Appendix 1. A MISAPPLICATION OF THE HARDY-WEINBERG LAW
167	Appendix 2. SOME POINTS IN STATISTICS
175	SUGGESTIONS FOR FURTHER READING
185	INDEX

In Memoriam—E. J. Machin
Best of Schoolmasters

Contents

	Page
Chapter 1. INTRODUCTION	1
Natural selection. Problems studied by population geneticists. Approaches to our problems. The Mendelian population. Factors affecting the genetic composition of populations. Summary.	
Chapter 2. SELFING	14
Inbreeding. Inbreeding depression. Difficulties in allowing for natural selection during inbreeding. Species inbreeding to some extent in nature. Effect of selfing on genotype frequencies at a single locus. Information needed to discuss the effect of inbreeding on the entire genotype. The start (generation 0). A later stage (generation n). Proportion of lines completely homozygous at a late stage. To determine the average number of tracts in a given generation. To determine the average number of heterozygous ends. To determine the average number of external junctions. Proportion of lines homozygous at all loci. An illustrative example. Total length heterozygous. Summary.	
Chapter 3. RANDOM GENETIC DRIFT	33
On the word "probability". Intuitive notions on the effects of drift. Probability distribution of allele frequency. Mean allele frequency: mean frequency of heterozygotes. Extreme versus intermediate allele frequencies. Circumstances under which drift may be ignored. Long-term effects of drift. Variable fertilities. Variable population size. Summary.	
Chapter 4. SURVIVAL AND FIXATION OF ADVANTAGEOUS MUTANTS	44
Factors affecting the chance of fixation. Genic selection. Mean and variance of the change in allele frequency. The probability of fixation: Kimura's formula. Accuracy of Kimura's formula. The probability of fixation: some conclusions. Initial frequency larger than $1/(2N)$. Variable population size (initial frequency $1/(2N)$). Dominance in fitness. Summary.	

Chapter 5.	WHAT IS THE ROLE OF DRIFT IN EVOLUTION? I. MORPHOLOGICAL CHARACTERS	72
	The problem. Harmful mutants. Obviously adaptive characters. "Minor" morphological variants. Chromosomal inversions in <i>Drosophila pseudoobscura</i> . Colour and banding in <i>Cepaea nemoralis</i> . Sternopleural chaetae in <i>Drosophila melanogaster</i> . Wing colour in <i>Panaxia dominula</i> . Conclusions on morphological characters.	
Chapter 6.	WHAT IS THE ROLE OF DRIFT IN EVOLUTION? II. PROTEIN VARIATION	87
	Protein polymorphism. Rate of nucleotide substitution: theory. Rate of nucleotide substitution in practice. A different approach. Allele frequencies in geographically distant areas. Effect of migration. Frequency of heterozygotes. Testing the neutral theory from frequency data: summary. Experimental approach. Experiments involving several loci. Predicting selective response from biochemical studies. Tentative conclusions. Summary.	
Chapter 7.	HITCH-HIKING EFFECT OF A FAVOURABLE GENE	109
	The hitch-hiking effect defined. Effect of recombination. The set-up most favourable for hitch-hiking. Some other situations leading to hitch-hiking. Summary.	
Chapter 8.	CHANGES OF ALLELE FREQUENCY UNDER NATURAL SELECTION IN LARGE RANDOM MATING POPULATIONS	115
	Conditions under which the problem can be simplified. Genic selection. Wright's formula. Change in allele frequency per generation (mild selection). Time required for a given change. Allowance for drift. Industrial melanism. Summary.	
Chapter 9.	EVOLUTION OF DOMINANCE	133
	Modifiers of dominance. Disadvantageous mutants. Advantageous mutants. Balanced polymorphism. Metrical characters. Summary.	

Chapter 10.	POLYMORPHISM	139
	Frequency of polymorphism. Some possible explanations for polymorphism. Heterozygous advantage: theory. Sickle-cell anaemia. Is heterozygous advantage common? Variable selective advantage: formal theory. Measuring selective advantage: formal approach. The ecological genetic approach. Inversions in <i>Drosophila pseudoobscura</i> . Batesian mimicry in butterflies. Final comments. Summary.	
Chapter 11.	EVOLUTION OF ALTRUISM	156
	Altruistic behaviour. Kin selection. Social insects. Group selection. Summary.	
Chapter 12.	SUMMARY. SOME TENTATIVE CONCLUSIONS	161
Appendix 1.	A MISAPPLICATION OF THE HARDY-WEINBERG LAW	164
	If genotype frequencies do not obey the Hardy-Weinberg law, what can we conclude? Problem. Solution. General comment.	
Appendix 2.	SOME POINTS IN STATISTICS	167
	Probability. Mean and variance. Probability distributions. Binomial distribution. Poisson distribution. Poisson approximation to binomial distribution. Continuous distributions. Normal distribution. Normal approximation to binomial distribution. A note on chapter 3. Some mathematical points.	
	SUGGESTIONS FOR FURTHER READING	175
	INDEX	185

CHAPTER ONE

INTRODUCTION

*But pardon, gentles all,
The flat unraised spirit, that hath dared,
On this unworthy scaffold to bring forth
So great an object...
...For the which supply,
Admit me chorus to this history;
Who, prologue-like, your humble patience pray,
Gently to hear, kindly to judge, our play.*

William Shakespeare, *King Henry V*

Natural selection

Hermann Bondi once wrote: "There can be no greater merit in a scientific discovery than that before long it should appear odd that it ever was considered a discovery... only those things that have so deeply affected our thinking and so thoroughly changed our outlook that we cannot think without them have really entered the spirit of the human race". These remarks, although made in another context, are entirely appropriate to the theory of evolution by natural selection. This theory is accepted (provisionally, as with any other scientific theory) because it conforms with the rest of our biological knowledge, because it explains the universal adaptation of living organisms to their environments, because specific examples of evolution by natural selection have been demonstrated in practice and, finally, because the theory can be formulated in a sufficiently precise way to be tested in practice. Thus the theory states that the process of natural selection is sufficient to bring about the changes that have taken place during the evolution of a particular character *in the time actually available*. Hence for a critical test of the notion that the evolution of some specific character has come

about through the agency of natural selection, we have first to show that the character is under genetic control, next to determine the selective advantage of the character to those individuals possessing it and, finally, to demonstrate that the magnitude of this advantage is large enough to have led to the changes that have actually occurred over a defined period of time. Critical tests of this kind can be carried out in cases of present-day evolution (although the practical difficulties are often formidable); thus the theory has the attribute "falsifiability in principle" generally held to be an essential feature of any useful scientific theory.

Problems studied by population geneticists

We shall centre all our later discussion on a number of problems critical to our understanding of the evolutionary process. Among these problems will be: Does natural selection within a group of organisms living in a particular habitat usually lead to uniformity or to diversity? What is the relationship between magnitude of selective advantages and rate of evolutionary change? Can natural selection give rise to self-sacrificing behaviour? How widespread are traits that convey some selective advantage?

Thus some workers, while accepting that obviously adaptive characters have spread through the agency of natural selection, consider that many other characters have spread purely by chance, it being a matter of indifference to the organism whether or not it carries the genes giving rise to such characters. This is the celebrated *neutral gene* theory. The opposing view is that natural selection is of near-universal occurrence and must be invoked in order to explain almost all the changes that have taken place in evolution. As we shall see, the neutralist and selectionist viewpoints lead to quite different predictions about natural populations, and many active attempts are being made to reach a decisive solution to this problem. It is quite wrong to suppose, as is occasionally done by those not familiar with the actual practice of population genetics, that any result not immediately explicable by natural selection is simply written off as a case where unknown forms of natural selection have operated. Readers of *Gulliver's Travels* will recall that such an appeal to occult causes was beneath the notice even of the imbecile scholars of the King of Brobdingnag. Selectionists, in common with other researchers, wish not to dismiss problems but to solve them; that is, to study problems, formulate hypotheses from which predictions can be made, and find out whether these predictions hold in practice.

Approaches to our problems

It might be supposed that the fossil record, so important for the elucidation of many aspects of evolution, would greatly assist the solution of the problems we have formulated. Unfortunately, much information essential to us is unavailable in the fossil record, which is therefore of limited help. Thus we do not know the magnitude of the selective advantage of successful types as compared with the unsuccessful types they replaced. We do not know the size of past interbreeding populations or the amount of inbreeding. However, we shall occasionally find the fossil record helpful; most useful in our context will be the study of rates of evolution. The fossil record suggests that evolution was a comparatively slow process in many cases. As we shall see, this observation in itself is compatible with either the neutralist or the selectionist theories, but an attempt to discriminate between these theories has been made by calculating for a given protein (e.g. haemoglobin) the average number of amino acid substitutions per amino acid site per year for a given line of descent (often called the rate of evolution of the protein for that line of descent). We shall discuss the details of the argument in chapter 6, but the basic idea is that if evolution of the protein was brought about by natural selection, the rate of evolution would vary from one line of descent to another. If, however, most of the amino acid substitutions were of no adaptive significance, it can be shown (given certain assumptions) that the rate would be about the same for different lines.

A much more useful approach, indicated briefly earlier, is the study of present-day natural populations. Here we can, at least in principle, estimate the magnitude of any factor (for example, population size, intensity of natural selection) relevant to the evolution of the population under investigation. However, such study of natural populations ("ecological genetics") is not always easy. One difficulty results from the gross heterogeneity of the natural habitat, a phenomenon which the reader can easily verify for himself. Frost cover, for example, may vary markedly over short distances. Thus natural selection may act quite differently in adjacent portions of the habitat, a genotype favoured in one portion being at a disadvantage in another. Similarly, the most successful genotype in a particular year might fail in another year when, say, climatic conditions were different.

In an effort to avoid these complications, experimenters often set up population cages in the laboratory, with a view to studying evolution

under more or less controlled conditions. The advantages of this approach are fairly obvious: different groups of individuals taken from a specific (natural or cage) population can be subjected to a variety of defined environmental variables, such as specified temperatures, humidity or food, and these treatments can be replicated so that the effects of other factors such as chance or uncontrolled environmental changes can be estimated. The main difficulty with cage experiments is that simplifying the natural habitat is, of course, falsifying it; for any individual experiment, we must try to decide whether the results would hold in nature. Despite this difficulty, cage experiments can be very helpful. To take a famous example (see chapter 10), suppose we observe that the frequencies of specific genotypes show marked fluctuations in the course of the year, the same fluctuations being found for several years in succession. We hypothesize that the relative selective advantages or disadvantages of the different genotypes fluctuate during the year as a result of changes in environmental factors such as food supply; we attempt, from such knowledge as we have of the biology of the organism concerned, to list such environmental factors. If we have correctly identified a relevant environmental factor, it should be possible to mimic the fluctuations in genotype frequency which are found in nature by varying this factor in cage experiments, an approach that has met with considerable success.

However, neither natural nor cage populations can in practice supply answers to all relevant problems. To take an example: it is often asked "can a selective advantage be so small that it does not matter?" More precisely, we might guess that if the selective advantage accruing to an individual from the possession of a particular allele is small, the allele will be "effectively neutral", that is, it will change in frequency in virtually the same way as does a neutral allele. We ask whether our guess is correct and if so how small must be the selective advantage for the allele to be effectively neutral. In principle, the question could be answered experimentally, by studying changes in frequency of alleles conveying small selective advantages. In practice, such an experiment is virtually impossible to carry out, since to estimate the selective advantage with the required degree of precision would involve an experiment on a truly enormous scale.

Thus there are points of great importance to us which cannot be investigated by observation or experiment. Hence we are forced to another approach—mathematical consideration of relatively simple models of evolution in action. Beginners often dislike such theoretical

studies. Insofar as this is not due merely to a distaste for mathematics as such (a distaste which we shall try to help such readers to overcome) the main objection appears to be as follows. Natural situations may well be very complicated, but the theory has not advanced sufficiently far to take account of most of these complications; the models used are gross over-simplifications of reality and so are liable to give very misleading conclusions. There is some justice in this objection. Obviously the answers we get will depend on the assumptions made, and these may sometimes be arbitrary or artificial.

Nevertheless, the objection is easily over-stated. Some results (e.g. that in large populations neutral alleles change in frequency very slowly) are almost certainly proof against any adjustments that might be made in the model. In the case of others, we can try adjusting the model in various ways and see what happens. We suggest that the reader approach the theory with a critical but open mind; ultimately, he must decide for himself the relevance of the results we shall give—theoretical, observational or experimental. It may help, however, at this preliminary stage, to recall an analogous situation in the physical sciences. When first attempting to study the flight of a projectile, we often ignore in the first instance the obviously relevant factor of air resistance; later, we may take account of the latter but ignore the rotation of the earth, although this is important in long-range gunnery. Yet the over-simplified models give useful results.

In summary, four main approaches are available: investigation of natural populations, of cage populations, of the fossil record and of theoretical models. We shall concentrate our discussions on specific problems, and for any such problem use any approach or approaches which seem appropriate for its resolution. For an alternative standpoint, taken by some authors, in which a particular approach is built up into a coherent body of knowledge with only occasional references to other approaches, the reader is referred to the literature (some suggestions for further reading are given at the end of this book).

The Mendelian population

In the interests of clarity, we define a Mendelian population as a group of inter-mating individuals. Thus an isolated group of 100 *strictly self-pollinating* plants living in the same area, although a population in the ecologist's sense, is not a single Mendelian population, size 100, but 100 Mendelian populations each of size one, since individuals never exchange

genes with other individuals. On the other hand, two groups of outbreeding plants which are geographically separated but which sometimes inter-mate by the passage of pollen from one group to the other, are all members of just one Mendelian population. However, for brevity's sake we shall (as hitherto) use the term *population* as short for Mendelian population, except in cases where confusion might arise.

We shall be concentrating subsequently on factors affecting the genetic composition of such a Mendelian population. Unless stated otherwise, we shall take it (for simplicity) that this population has constant size N and that different generations do not overlap; usually, individuals will be diploid. We shall often consider just one locus at a time, with just two alleles at a locus. If at some time there are NP individuals of genotype AA, $2NQ$ individuals of genotype Aa and NR individuals of genotype aa (the use of $2NQ$ rather than NQ for Aa will simplify later calculations) then the proportions of these three genotypes (the "genotype frequencies") at that time are

Genotype	<u>AA</u>	<u>Aa</u>	<u>aa</u>	Total
Frequency	P	$2Q$	R	1

the total being unity since we have assumed only two alleles present in the population. These genotype frequencies summarize the genetic composition of the population at the particular time. Since the frequencies add to unity, it would be sufficient to give two of them, the remaining frequency being immediately obtainable by subtraction.

A simpler summary is given by the frequencies of the two alleles (the *allele frequencies*—often referred to, slightly loosely, as the "gene frequencies"). Since individuals are diploid, there will be $2N$ alleles at the locus in the population as a whole; of these alleles ($2NP + 2NQ$) will be A, ($2NQ + 2NR$) will be a, so that the allele frequencies are

Allele	<u>A</u>	<u>a</u>	Total
Frequency	$P + Q = p$ (say)	$Q + R = q$ (say)	1

Since $p + q = 1$, it is sufficient to give either p or q . Whenever possible, we shall give the simpler summary in terms of one allele frequency (usually p) rather than the more complicated summary in terms of two genotype frequencies.

Factors affecting the genetic composition of populations

The evolutionary process is often envisaged in terms of changes in allele frequencies or in genotype frequencies; the multiplicity of phenomena bringing about changes in genotype frequency are grouped under the headings: mating system, chance, natural selection, mutation. If, as often happens, the population is divided into sub-populations, with members of any sub-population breeding freely *inter se* but in a more restricted way with members of other sub-populations, then different sub-populations may differ from one another in their genetic composition. In such a case, an additional phenomenon, the migration of individuals or gametes (e.g. pollen) from one sub-population to another, can have an important effect on frequencies.

It is easiest, in the first instance, to consider each group of phenomena on its own, ignoring for the moment the effects of the others.

A fundamental consequence of Mendelian inheritance may be stated informally as: the hereditary mechanism, of itself, does not change the allelic composition of a population. More formally we say: in the absence of chance effects, natural selection and mutation, the allele frequencies remain constant over generations, whatever the mating system; for, with no chance effects, selection or mutation, genotype frequencies and hence allele frequencies p, q remain unchanged throughout the life cycle. Now (still with the same preconditions) consider gametogenesis: the proportion of gametes containing A will be p , the proportion of gametes containing a will be q , that is, gametic frequencies equal allele frequencies. Irrespective of the way in which these gametes are combined into zygotes giving rise to the next generation, allele frequencies in this next generation will remain unchanged at values p, q .

On the other hand, the mating system will determine how gametes are combined into zygotes, and will thus affect zygotic frequencies, as we shall show shortly by consideration of two different mating systems. Some examples of different mating systems are: random mating (the chance that an individual mates with a particular genotype equals the frequency of that genotype in the population); positive assortative mating (like phenotypes tend to mate); inbreeding (close relatives tend to mate)—we consider this more fully in chapter 2. Note that the mating system is not necessarily the same for different loci. For example, when choosing a mate, a human being may note many aspects of the potential partner's phenotype, but hardly their blood group. Not surprisingly,

then, we find in our species positive assortative mating for stature, whereas blood groups provide excellent examples of random mating.

Under random mating, genotype frequencies among zygotes are very simply related to the frequencies of the gametes from which these zygotes were derived. Suppose gametic frequencies are \underline{A} p , \underline{a} q . Now random mating implies random union of gametes (for a discussion of this point, see Edwards 1977). Thus in the absence of mutation, selection on the gametes or chance effects, we have

Type of union		Frequency	Derived zygote
Egg	Sperm		
\underline{A}	\underline{A}	p^2	\underline{AA}
\underline{A}	\underline{a}	pq	\underline{Aa}
\underline{a}	\underline{A}	qp	\underline{Aa}
\underline{a}	\underline{a}	q^2	\underline{aa}

so that frequencies of the zygotes are

$$\underline{AA} \quad p^2, \quad \underline{Aa} \quad 2pq, \quad \underline{aa} \quad q^2$$

The new allele frequencies are

$$\underline{A} \quad p^2 + pq = p(p + q) = p$$

$$\underline{a} \quad q^2 + pq = q(q + p) = q$$

since $p + q = 1$. Thus the new allele frequencies are equal to the gametic frequencies, and hence to the allele frequencies one generation preceding, as we have already established earlier for the more general case.

With the allele frequencies unchanged, the new gametic frequencies will be the same as those one generation preceding. It is apparent, then, that genotype frequencies, established in one generation at p^2 , $2pq$, q^2 remain at these values indefinitely. This result, known as the Hardy-Weinberg law, may be stated in the form

“a sufficient condition for no evolution to occur within a Mendelian population is that mutation, selection and chance effects are all absent and that mating is at random”.

If we wish to apply the law to a population in the ecologist's sense, we must have the additional precondition “no migration”.

The Hardy-Weinberg law is best regarded as a special case of the fundamental principle, given earlier, that the hereditary mechanism, of itself, does not change allele frequencies. The constancy of the genotypic